
Niveau(x) d'enseignement: 4
Centre d'intérêt: 840

Numéro EDUQ : 9370177

Arsenault, Louise.

Sciences : atelier de perfectionnement en évaluation sommative à partir du modèle retenu dans les définitions du domaine : enseignement secondaire, 4e et 5e année : recueil d'activités / [conception et rédaction, Louise Arsenault]. -- [Québec] : Direction générale de l'évaluation et des ressources didactiques, 1992.

236 f. en foliotation multiple : ill. ; 28 cm.

Code: 9192-0830; 16-7560.

Vedette(s)-secondaire(s): Québec (Province). Direction générale de l'évaluation et des ressources didactiques

Descripteur(s): Sciences * Programme d'études * Evaluation sommative * Evaluation de l'apprentissage * Instrument de mesure * Elaboration d'examens * Validité * Correction * Evaluation de tests * Perfectionnement des enseignants * Secondaire (Ordre d'enseignement)

Identificateur(s) de sujet: Fidélité

L'atelier de perfectionnement a pour objectif de rendre les participants habiles à utiliser les définitions du domaine proposées dans les programmes d'études afin qu'ils puissent bâtir des épreuves de qualité. Six activités font partie de l'atelier: 1- les fondements de la mesure et de l'évaluation des apprentissages; 2- les qualités des instruments de mesure; 3- les définitions du domaine en sciences; 4- la construction d'épreuves répondant aux exigences des qualités de validité et de fidélité; 5- la correction de l'épreuve; 6- la qualité des items d'une épreuve jugée à l'aide des indices de difficulté et de discrimination. Un guide de l'animateur et un guide du participant accompagnent chacune des activités. Un document de base donnant des précisions sur les définitions du domaine est présenté à la suite des activités.

Type de document:

9370177

SCIENCES

*Atelier de perfectionnement
en évaluation sommative*

*À partir du modèle retenu
dans les définitions du domaine*

*Enseignement secondaire
4^e et 5^e année*

RECUEIL D'ACTIVITÉS

Québec ::

9370177

SCIENCES

*Atelier de perfectionnement
en évaluation sommative*

*À partir du modèle retenu
dans les définitions du domaine*

*Enseignement secondaire
4^e et 5^e année*

RECUEIL D'ACTIVITÉS

Afin de faciliter la lecture du présent document, les rédacteurs et les rédactrices ont considéré le masculin comme générique ou neutre. Lecteurs et lectrices comprendront que cet emploi est fait sans porter préjudice à quiconque.

© Gouvernement du Québec
Ministère de l'Éducation, 1992-9192-0830

ISBN 2-550-23013-2

Dépôt légal - premier trimestre 1992
Bibliothèque nationale de Québec

SCIENCES

*Atelier de perfectionnement
en évaluation sommative*

*À partir du modèle retenu
dans les définitions du domaine*

*Enseignement secondaire
4^e et 5^e année*

RECUEIL D'ACTIVITÉS

COORDINATION

Mihran Djiknavorian

Direction du développement de l'évaluation

CONCEPTION ET RÉDACTION

Louise Arsenault

Consultante en mesure et évaluation

RÉVISION LINGUISTIQUE

Sous la responsabilité du ministère de l'Éducation

CONSULTATION

Diane Caouette-Tardif

CSR de Chambly

Michel Charette

CS Aylmer

Jean-Marie Demers

CS Rouyn-Noranda

Albert Dolan

CS de Sept-Îles

Jocelyn Drouin

CS des Découvreurs

André Flibotte

CS Saint-Jérôme

Réginal Fortier

CS de l'Asbesterie

Jacques Gélinas

CS Samuel-De Champlain

Jacques Girard

CS de Manicouagan

Éva Grenier

CSR Louis-Frédette

Guy Lapointe

CS d'Iberville

Louisianna Lévesque

CS des Draveurs

Gilles Michaud

CSR Chauveau

Marie-Ange Soucy

CS La Neigette

Michel Tremblay

CS de Dolbeau

Margot Vachon

CS Jérôme-Le Royer

CONTENU DU RECUEIL D'ACTIVITÉS

OBJECTIF DE L'ATELIER

LES ACTIVITÉS

ACTIVITÉ 1

Les fondements de la mesure et de l'évaluation des apprentissages

Guide de l'animateur

Guide du participant

ACTIVITÉ 2

Les qualités des instruments de mesure

Guide de l'animateur

Guide du participant

ACTIVITÉ 3

Les définitions du domaine en sciences

Guide de l'animateur

Guide du participant

ACTIVITÉ 4

La construction d'épreuves répondant aux exigences des qualités de validité et de fidélité

Guide de l'animateur

Guide du participant

ACTIVITÉ 5

La correction de l'épreuve

Guide de l'animateur

Guide du participant

ACTIVITÉ 6

La qualité des items d'une épreuve jugée à l'aide des indices de difficulté et de discrimination

Guide de l'animateur

Guide du participant

DOCUMENT DE BASE

FORMULES D'ÉVALUATION DE L'ATELIER

OBJECTIF DE L'ATELIER

L'atelier a pour objectif de rendre les participants habiles à utiliser les définitions du domaine proposées dans les programmes d'études afin qu'ils puissent bâtir des épreuves de qualité.

LES ACTIVITÉS

ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 1

Les fondements de la mesure et de l'évaluation des apprentissages

GUIDE DE L'ANIMATEUR

ACTIVITÉ 1

Les fondements de la mesure et de l'évaluation des apprentissages

1. *Les limites de la mesure*
2. *La mesure, l'évaluation et la décision*
3. *L'évaluation sommative et l'évaluation formative*
4. *L'interprétation normative et l'interprétation critérielle*

1. LES LIMITES DE LA MESURE

Chaque fois que nous prenons une mesure, nous obtenons un score. Ce score peut être affecté d'erreurs de mesure.

Exemple : la température.

Si je mesure la température de l'air, j'obtiens un certain degré qui peut être affecté d'une erreur de mesure.

Qu'est-ce qui peut causer cette erreur de mesure?

Réponses possibles :

Mon instrument, le thermomètre.

Ma vue, la lecture de l'instrument.

L'endroit précis où je me situe lorsque je prends la mesure (courant d'air, près d'un calorifère...).

Lorsque je mesure le rendement des élèves en sciences, il se produit le même phénomène. J'obtiens un «score» qui peut être affecté d'erreurs de mesure. L'équation suivante résume bien le phénomène.

ACÉTATE 1

$$V = X \pm E$$

$V =$ Le «vrai» score de l'élève.

$X =$ Le résultat de l'élève à une épreuve.

$E =$ L'erreur de mesure qui peut être positive ou négative.

Donc, lorsque je mesure le rendement de mes élèves en sciences, qu'est-ce qui peut causer une erreur de mesure?

Donnez-moi des sources d'erreurs possibles.

Par la suite, résumez les interventions des participants en précisant les trois grandes sources d'erreurs.

Erreurs provenant de l'instrument :

- item ambigu,
- item non congruent avec l'objectif mesuré,
- longueur de l'épreuve,
- vocabulaire utilisé,
- temps limité,
- correction, etc.

Erreurs provenant des conditions d'administration de l'épreuve :

- moment,
- conditions non uniformes pour tous les élèves, etc.

Erreurs provenant de l'élève :

- motivation,
- apprentissage et éducation,
- expérience des épreuves,
- nervosité,
- variables physiologiques,
- intelligence,
- habileté à lire,
- habileté à travailler rapidement, etc.

De plus, certains aspects comme la résolution de problèmes, les objectifs affectifs et les attitudes ne s'évaluent pas directement. Ce sont les traces de ces phénomènes que nous observons. La mesure prise est plutôt indirecte, ce qui constitue une autre source d'erreurs.

La mesure de l'habileté à résoudre un problème :

le début _____ la fin
du problème (démarche) du
problème

De tout ce que l'élève fait ou pense durant la résolution du problème (la démarche), on ne connaît que les traces qu'il laisse : orales ou écrites, graphiques ou de comportement... Ceci constitue une observation indirecte de l'habileté à résoudre des problèmes. En évaluation formative, cette limite ne représente pas une difficulté en soi.

En évaluation sommative, au moment de la sanction des études, nous n'avons que les traces sur la copie de l'élève pour accorder des points à la démarche.

Il faut donc habituer l'élève à décrire sa démarche, l'habituer même à décrire l'endroit où il bloque!

Enfin, même en prenant le plus grand soin, nous ne pourrons jamais affirmer que la mesure que nous obtenons est exempte de toute erreur en raison de l'élève lui-même. Nous n'avons aucun contrôle sur cette source d'erreurs.

Somme toute, vous êtes probablement les plus sensibilisés aux limites de la mesure.

Vous êtes les personnes qui jouez le plus avec ces facteurs, vous les connaissez mieux que personne. Vous connaissez aussi les répercussions des limites de la mesure.

Pourquoi vous en parle-t-on?

Parfois on ressent un malaise au regard de l'évaluation de nos élèves.

Ce malaise s'explique souvent par la nature même de la tâche qui nous est demandée (équation).

Le laboratoire où vous prenez des mesures est grand...

Nous allons donc tenter durant l'atelier de circonscrire les endroits où il nous est possible d'intervenir afin de minimiser l'erreur de mesure.

2. LA MESURE, L'ÉVALUATION ET LA DÉCISION

Quelle est la différence entre une mesure et une évaluation?

Mesurer : opération qui consiste à assigner des nombres à des objets, à des personnes ou à des événements en suivant une règle précise. C'est l'opération qui conduit à la mesure.

La mesure est simplement un chiffre, un nombre ou une quantité.

L'évaluation est le jugement, l'interprétation que nous allons faire de la mesure.

Exemple :

8/10 : À quoi correspond cette valeur?
Quelle est cette valeur?

C'est une mesure.

Justine a obtenu 8/10 à une épreuve sommative en sciences à la fin de l'année scolaire.

Que pouvez-vous dire de plus?

C'est un bon rendement...

C'est bien pour Justine...

Elle aurait pu faire mieux...

Vous venez d'interpréter, de porter un jugement, d'évaluer le résultat de Justine.

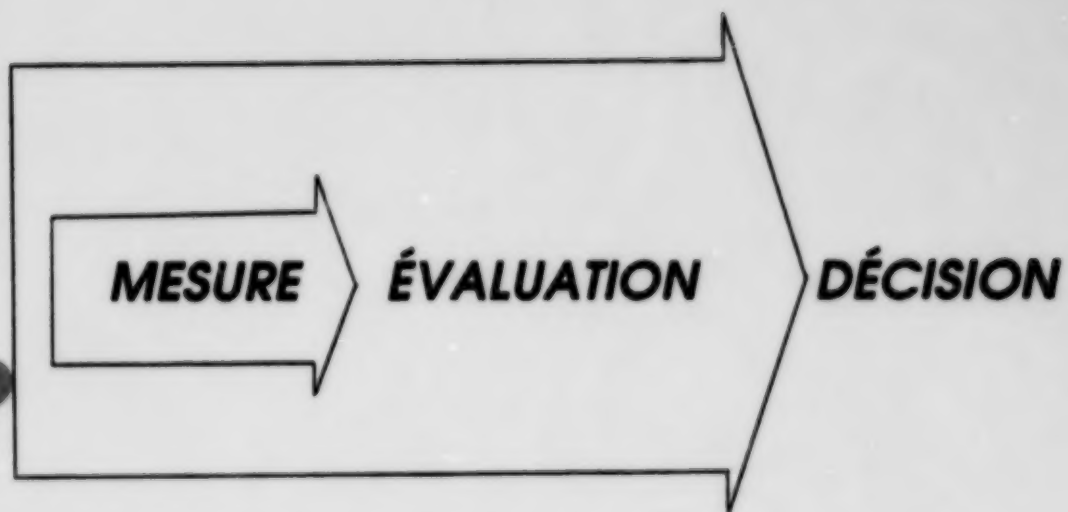
Quelle décision pourriez-vous prendre suite à cette constatation?

Justine a réussi le programme de sciences, elle peut passer à une classe supérieure.

La version contemporaine de l'évaluation intègre la notion de mesure comme étape préalable dans le processus même d'évaluation. On reconnaît également que le processus d'évaluation a une destination : la prise de décision.

Il n'en fut pas toujours ainsi.

ACÉTATE 2



Un exercice : Mesure ou évaluation?

ACÉTATE 3

MESURE ET ÉVALUATION

Pour chacun des énoncés suivants, indiquez (M) si c'est un résultat de mesure (une observation factuelle, une constatation) ou indiquez (E) si c'est une évaluation (un résultat interprété, un jugement).

- | | | |
|-----|---|-------|
| 1. | Une grande maison. | () |
| 2. | Une personne âgée de 42 ans. | () |
| 3. | Une grosse tempête de neige. | () |
| 4. | Un angle de 38 degrés. | () |
| 5. | Une personne agile. | () |
| 6. | Un grand bureau. | () |
| 7. | Une note de 30 points à 30 questions. | () |
| 8. | Un bon professeur. | () |
| 9. | Un rendement moyen. | () |
| 10. | Un phénomène se produisant trois fois par mois. | () |
| 11. | Un résultat de 78 %. | () |
| 12. | Un atelier d'une longue durée. | () |
| 13. | Des participants charmants. | () |
| 14. | Huit problèmes réussis sur dix. | () |
| 15. | Une pause de 20 minutes. | () |

MESURE ET ÉVALUATION

Pour chacun des énoncés suivants, indiquez (M) si c'est un résultat de mesure (une observation factuelle, une constatation) ou indiquez (E) si c'est une évaluation (un résultat interprété, un jugement).

- | | | |
|-----|---|-------|
| 1. | Une grande maison. | (E) |
| 2. | Une personne âgée de 42 ans. | (M) |
| 3. | Une grosse tempête de neige. | (E) |
| 4. | Un angle de 38 degrés. | (M) |
| 5. | Une personne agile. | (E) |
| 6. | Un grand bureau. | (E) |
| 7. | Une note de 30 points à 30 questions. | (M) |
| 8. | Un bon professeur. | (E) |
| 9. | Un rendement moyen. | (E) |
| 10. | Un phénomène se produisant trois fois par mois. | (M) |
| 11. | Un résultat de 78 %. | (M) |
| 12. | Un atelier d'une longue durée. | (E) |
| 13. | Des participants charmants. | (E) |
| 14. | Huit problèmes réussis sur dix. | (M) |
| 15. | Une pause de 20 minutes. | (M) |

3. L'ÉVALUATION SOMMATIVE ET L'ÉVALUATION FORMATIVE

L'évaluation sommative est une «démarche qui vise à porter un jugement sur le degré de réalisation des apprentissages visés par un programme ou une partie terminale de programme ou encore d'un ensemble équivalent d'apprentissages faits en dehors de l'école, en se basant sur des données pertinentes, pour prendre des décisions relatives :

- au passage à la classe supérieure,
- à la sanction des études,
- à la reconnaissance des acquis expérimentiels admis en équivalences,
- à l'orientation de l'élève ¹»

En ce qui nous concerne, le type de décision à prendre est surtout relatif à la sanction des études, à la délivrance des diplômes.

Autrement dit, l'évaluation sommative sert à informer l'élève et l'enseignant sur la maîtrise d'un ensemble d'objectifs du programme d'études pour faciliter la prise de décision relativement à la sanction des apprentissages.

Sanctionner est différent de sélectionner.

Sanctionner est différent de classer ou ordonner.

Pour qu'elle soit sommative (décision relative à la sanction des apprentissages), l'évaluation doit répondre à des exigences bien particulières et ses scénarios sont plutôt rigides, ce qui la distingue très nettement de l'évaluation formative.

Si le mandat est de sélectionner, classer ou ordonner, on doit imaginer des scénarios très différents de ceux utilisés pour sanctionner des apprentissages.

ACÉTATE 4

1. MEQ (Ministère de l'Éducation du Québec), *Éléments de docimologie, lexique*, fascicule 2, 1985, p. 10, 11.

Enfin, lorsque l'on imagine des scénarios en évaluation sommative, ces derniers doivent être rigides car on souhaite une base semblable pour certifier les apprentissages d'un même programme d'études.

Pour l'enseignant qui prépare une épreuve destinée aux élèves de sa classe, c'est relativement simple de les soumettre tous aux mêmes conditions. La marge de manoeuvre est grande.

Par contre, lorsqu'une équipe élabore une épreuve à l'échelle provinciale, des contraintes s'ajoutent. La marge de manoeuvre est beaucoup moins grande. À ce moment, les définitions du domaine vont aider à donner une base semblable à l'évaluation des objectifs d'un programme d'études et vont en faciliter la communication à tous les enseignants intéressés. La définition du domaine devient donc un préalable lorsqu'on désire soumettre plusieurs groupes d'élèves à des conditions identiques.

Ces préoccupations ne se posent pas en évaluation formative.

L'évaluation formative, quant à elle, est «un processus d'évaluation continue ayant pour objet d'assurer la progression de chaque individu dans une démarche d'apprentissage avec l'intention de modifier le rythme de cette progression ou la démarche elle-même, pour apporter (s'il y a lieu) des améliorations ou des correctifs appropriés²».

Autrement dit, l'évaluation formative est composée de tous les moyens que vous prenez pour aider l'élève à apprendre pendant qu'il apprend. Ces moyens peuvent être différents d'un élève à l'autre puisque ces derniers sont différents.

2. Gérard Scallon, *Définition, exigences et caractéristiques de l'évaluation formative*, notes de cours.

En évaluation formative, tout peut être évalué, même ce qui n'est pas prescrit dans les objectifs du programme d'études, à la condition que ce soit fait pendant la période «enseignement et apprentissage» et que vous ayez l'intention de favoriser l'apprentissage d'un ou de plusieurs élèves.

L'évaluation formative a donc pour cible des apprentissages inachevés, ce qui exclut la consignation des résultats dans le bulletin de l'élève.

L'évaluation formative n'a pas un caractère de sanction des apprentissages, elle s'inscrit dans une relation d'aide.

La participation de l'élève peut être sollicitée au moment de l'évaluation formative.

Il n'est pas nécessaire d'utiliser des instruments de mesure pour l'évaluation formative. L'observation peut être un excellent moyen de donner une rétroaction à l'élève afin de favoriser son apprentissage.

En évaluation formative, les notes (résultat de mesure) ne sont pas indispensables.

En évaluation formative, la mesure des apprentissages visés suppose la plupart du temps un caractère analytique, la note totale présente donc peu d'intérêt.

En évaluation formative, ce n'est pas toujours l'élève qui est jugé; ce peut être votre enseignement, du matériel pédagogique, etc.

L'acétate 4 présente un parallèle entre l'évaluation sommative et l'évaluation formative.

ÉVALUATION SOMMATIVE ET ÉVALUATION FORMATIVE

	<u>Évaluation sommative</u>	<u>Évaluation formative</u>
<i>Intention de l'enseignant</i>	<i>Sanctionner des apprentissages.</i>	<i>Favoriser la progression de l'apprentissage.</i>
<i>Objet de l'évaluation</i>	<i>Les objectifs terminaux du programme d'études (apprentissage terminé).</i>	<i>Des préalables, la démarche de l'élève, des apprentissages inachevés ou modifiables, des stratégies.</i>
<i>Scénario</i>	<i>Rigide, semblable pour tous les élèves soumis à des conditions identiques.</i>	<i>Souple, parfois différent d'un élève à l'autre. La participation de l'élève peut être sollicitée.</i>
<i>Moment</i>	<i>À la fin d'une période d'enseignement et d'apprentissage.</i>	<i>Pendant la période d'enseignement et d'apprentissage.</i>

Un exercice : Évaluation formative ou évaluation
sommativ

ACÉTATE 5

Références :

MEQ, *Éléments de docimologie, l'évaluation formative*,
fascicule 3, 1985, 118 p.

MEQ, *Éléments de docimologie, l'évaluation sommativ*,
fascicule 4, 1987, 89 p.

MEQ, *Politique générale d'évaluation pédagogique*,
Secteur du préscolaire, du primaire et du secondaire,
1981, 23 p.

ÉVALUATION FORMATIVE ET ÉVALUATION SOMMATIVE

1-5

Pour chacun des énoncés suivants, indiquez (EF) s'il appartient à l'évaluation formative ou indiquez (ES) s'il appartient à l'évaluation sommative.

1. Un élève éprouve certaines difficultés en sciences; son professeur le rencontre pour sonder son degré de motivation à l'étude. ()
2. Une épreuve de sciences est administrée à la fin d'une étape; les scores sont convertis en rangs pour être consignés dans un bulletin. ()
3. Les résultats de l'épreuve de juin 1991 sont utilisés en septembre 1991 pour préparer des activités de révision.
(Répondre en fonction des élèves qui ont subi l'épreuve.) ()
4. Un enseignant interroge oralement ses élèves afin de décider s'il reviendra sur certaines notions. ()
5. Une enseignante administre une épreuve dont la cible sera des apprentissages inachevés. ()
6. Dans le but de sanctionner des apprentissages, une épreuve portant sur l'ensemble des objectifs du programme de sciences est administrée à la fin de l'année. ()
7. Dans le but de favoriser la progression de l'apprentissage, une épreuve portant sur un seul objectif est administrée pendant que les élèves réalisent cet apprentissage. ()
8. Les élèves subissent, à chaque semaine, une épreuve de rendement; les scores de ces épreuves sont additionnés de façon à produire pour chaque élève un bilan qui servira d'évaluation à la fin du cours. ()

ÉVALUATION FORMATIVE ET ÉVALUATION SOMMATIVE

Pour chacun des énoncés suivants, indiquez (EF) s'il appartient à l'évaluation formative ou indiquez (ES) s'il appartient à l'évaluation sommative.

1. Un élève éprouve certaines difficultés en sciences; son professeur le rencontre pour sonder son degré de motivation à l'étude. (EF)
2. Une épreuve de sciences est administrée à la fin d'une étape; les scores sont convertis en rangs pour être consignés dans un bulletin. (ES)
3. Les résultats de l'épreuve de juin 1991 sont utilisés en septembre 1991 pour préparer des activités de révision.
(Répondre en fonction des élèves qui ont subi l'épreuve.) (ES)
4. Un enseignant interroge oralement ses élèves afin de décider s'il reviendra sur certaines notions. (EF)
5. Une enseignante administre une épreuve dont la cible sera des apprentissages inachevés. (EF)
6. Dans le but de sanctionner des apprentissages, une épreuve portant sur l'ensemble des objectifs du programme de sciences est administrée à la fin de l'année. (ES)
7. Dans le but de favoriser la progression de l'apprentissage, une épreuve portant sur un seul objectif est administrée pendant que les élèves réalisent cet apprentissage. (EF)
8. Les élèves subissent, à chaque semaine, une épreuve de rendement; les scores de ces épreuves sont additionnés de façon à produire pour chaque élève un bilan qui servira d'évaluation à la fin du cours. (ES)

4. L'INTERPRÉTATION NORMATIVE ET L'INTERPRÉTATION CRITÉRIELLE

Interpréter un score de façon critérielle consiste essentiellement à comparer ce score à un critère fixé à l'avance et pris comme seuil de réussite.

Interpréter un score de façon normative consiste essentiellement à comparer ce score à une norme (moyenne de groupe).

En évaluation sommative, on peut interpréter les résultats de façon normative ou critérielle.

En évaluation formative, on interprète les résultats de façon critérielle.

Un exercice : *Interprétation normative ou interprétation critérielle?*

ACÉTATE 6

Références :

MEQ, *Éléments de docimologie, l'évaluation pédagogique : une démarche*, fascicule 1, 1985, 29 p.

MEQ, *Éléments de docimologie, lexique*, fascicule 2, 1985, 23 p.

MEQ, *Éléments de docimologie, l'évaluation formative*, fascicule 3, 1985, 118 p.

MEQ, *Éléments de docimologie, l'évaluation sommative*, fascicule 4, 1987, 89 p.

MEQ, *Politique générale d'évaluation pédagogique, Secteur du préscolaire, du primaire et du secondaire*, 1981, 23 p.

INTERPRÉTATION NORMATIVE OU INTERPRÉTATION CRITÉRIELLE

Pour chacun des énoncés suivants, indiquez (IN) si les résultats sont interprétés de façon normative ou indiquez (IC) si les résultats sont interprétés de façon critérielle.

1. À une compétition de ski, on remet une médaille aux trois personnes qui ont été les premières à franchir la ligne d'arrivée. ()
2. À la compétition des BONS VIVANTS, on remet un prix à tous les skieurs qui ont pris moins de cinq minutes à franchir la ligne d'arrivée et ce, sans tomber. ()
3. À une épreuve de sciences, seuls les élèves qui auront réussi huit problèmes sur dix pourront aborder le chapitre suivant. ()
4. À l'épreuve de juin en sciences, les dix élèves les plus forts obtiendront la lettre A. ()
5. Comparer les élèves entre eux pour choisir le plus grand. ()
6. D'après un règlement du ministère de l'Éducation, seuls les enfants ayant atteint l'âge prescrit au 30 septembre seront autorisés à commencer leurs études primaires. ()
7. Pour interpréter le résultat d'un élève à une épreuve sommative, son score est transformé en rang stanine. ()

INTERPRÉTATION NORMATIVE OU INTERPRÉTATION CRITÉRIELLE

Pour chacun des énoncés suivants, indiquez (IN) si les résultats sont interprétés de façon normative ou indiquez (IC) si les résultats sont interprétés de façon critérielle.

1. À une compétition de ski, on remet une médaille aux trois personnes qui ont été les premières à franchir la ligne d'arrivée. (IN)
2. À la compétition des BONS VIVANTS, on remet un prix à tous les skieurs qui ont pris moins de cinq minutes à franchir la ligne d'arrivée et ce, sans tomber. (IC)
3. À une épreuve de sciences, seuls les élèves qui auront réussi huit problèmes sur dix pourront aborder le chapitre suivant. (IC)
4. À l'épreuve de juin en sciences, les dix élèves les plus forts obtiendront la lettre A. (IN)
5. Comparer les élèves entre eux pour choisir le plus grand. (IN)
6. D'après un règlement du ministère de l'Éducation, seuls les enfants ayant atteint l'âge prescrit au 30 septembre seront autorisés à commencer leurs études primaires. (IC)
7. Pour interpréter le résultat d'un élève à une épreuve sommative, son score est transformé en rang stanine. (IN)



**ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE**

ACTIVITÉ 1

Les fondements de la mesure et de l'évaluation des apprentissages

GUIDE DU PARTICIPANT

ACTIVITÉ 1

Les fondements de la mesure et de l'évaluation des apprentissages

1. *Les limites de la mesure*
2. *La mesure, l'évaluation et la décision*
3. *L'évaluation sommative et l'évaluation formative*
4. *L'interprétation normative et l'interprétation critérielle*

1. LES LIMITES DE LA MESURE

Chaque fois que nous prenons une mesure, nous obtenons un score. Ce score peut être affecté d'erreurs de mesure.

Exemple : la température.

Si je mesure la température de l'air, j'obtiens un certain degré qui peut être affecté d'une erreur de mesure.

Qu'est-ce qui peut causer cette erreur de mesure?

Réponses possibles :



Lorsque je mesure le rendement des élèves en sciences, il se produit le même phénomène. J'obtiens un «score» qui peut être affecté d'erreurs de mesure. L'équation suivante résume bien le phénomène.

$$V = X \pm E$$

V = Le «vrai» score de l'élève.

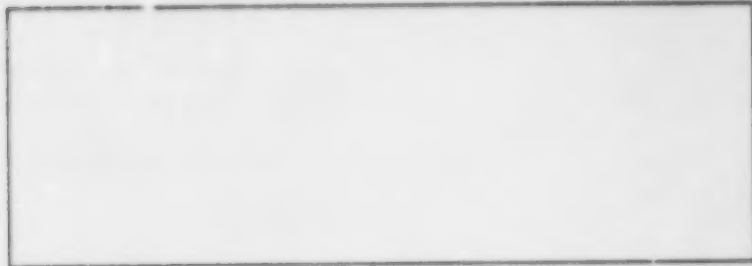
X = Le résultat de l'élève à une épreuve.

E = L'erreur de mesure qui peut être positive ou négative.

Donc, lorsque je mesure le rendement de mes élèves en sciences, qu'est-ce qui peut causer une erreur de mesure?

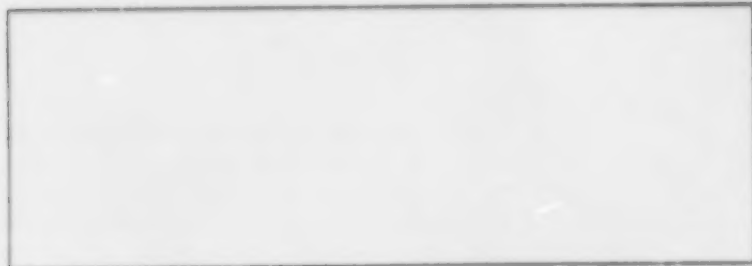
Donnez-moi des sources d'erreurs possibles.

Réponses possibles :

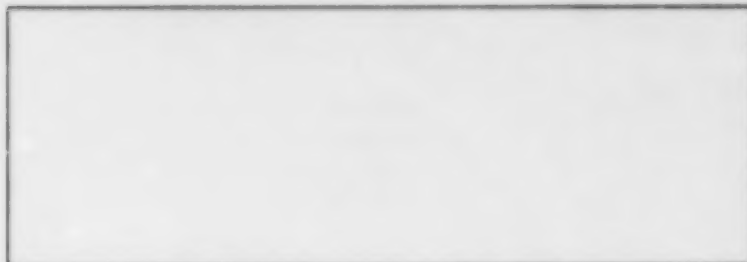


Les trois grandes sources d'erreurs

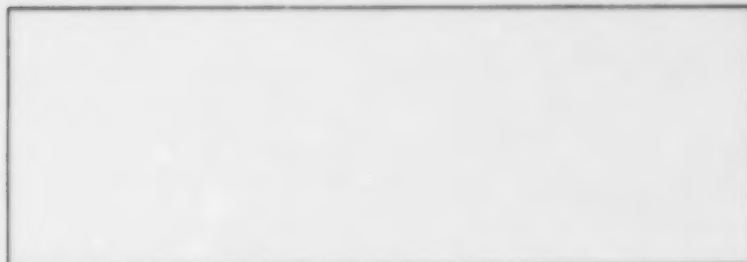
Erreurs provenant de l'instrument :



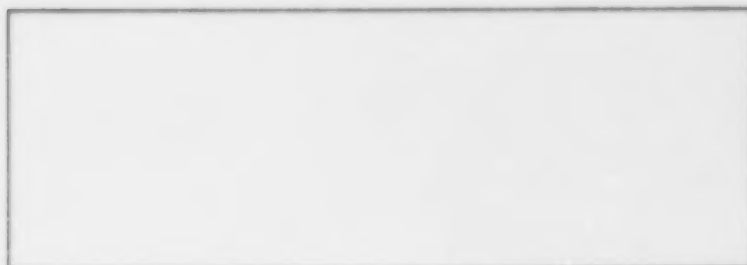
*Erreurs provenant des conditions d'administration de
l'épreuve :*



Erreurs provenant de l'élève :



Autres limites :



2. LA MESURE, L'ÉVALUATION ET LA DÉCISION

Quelle est la différence entre une mesure et une évaluation?

Mesurer : opération qui consiste à assigner des nombres à des objets, à des personnes ou à des événements en suivant une règle précise. C'est l'opération qui conduit à la mesure.

La mesure est simplement un chiffre, un nombre ou une quantité.

L'évaluation est le jugement, l'interprétation que nous allons faire de la mesure.

Exemple :

8/10 : À quoi correspond cette valeur?
Quelle est cette valeur?

Justine a obtenu 8/10 à une épreuve sommative en sciences à la fin de l'année scolaire.

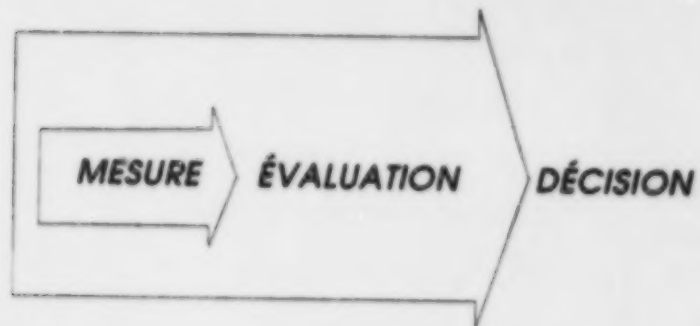
Que pouvez-vous dire de plus?

Vous venez d'interpréter, de porter un jugement, d'évaluer le résultat de Justine.

Quelle décision pourriez-vous prendre suite à cette constatation?

La version contemporaine de l'évaluation intègre la notion de mesure comme étape préalable dans le processus même d'évaluation. On reconnaît également que le processus d'évaluation a une destination: la prise de décision.

Il n'en fut pas toujours ainsi.



MESURE ET ÉVALUATION

Pour chacun des énoncés suivants, indiquez (M) si c'est un résultat de mesure (une observation factuelle, une constatation) ou indiquez (E) si c'est une évaluation (un résultat interprété, un jugement).

- | | | |
|-----|---|-------|
| 1. | Une grande maison. | () |
| 2. | Une personne âgée de 42 ans. | () |
| 3. | Une grosse tempête de neige. | () |
| 4. | Un angle de 38 degrés. | () |
| 5. | Une personne agile. | () |
| 6. | Un grand bureau. | () |
| 7. | Une note de 30 points à 30 questions. | () |
| 8. | Un bon professeur. | () |
| 9. | Un rendement moyen. | () |
| 10. | Un phénomène se produisant trois fois par mois. | () |
| 11. | Un résultat de 78 %. | () |
| 12. | Un atelier d'une longue durée. | () |
| 13. | Des participants charmants. | () |
| 14. | Huit problèmes réussis sur dix. | () |
| 15. | Une pause de 20 minutes. | () |

3. L'ÉVALUATION SOMMATIVE ET L'ÉVALUATION FORMATIVE

L'évaluation sommative est une «démarche qui vise à porter un jugement sur le degré de réalisation des apprentissages visés par un programme ou une partie terminale de programme ou encore d'un ensemble équivalent d'apprentissages faits en dehors de l'école, en se basant sur des données pertinentes, pour prendre des décisions relatives :

- au passage à la classe supérieure,
- à la sanction des études,
- à la reconnaissance des acquis expérimentiels admis en équivalences,
- à l'orientation de l'élève ¹»

En ce qui nous concerne, le type de décision à prendre est surtout relatif à la sanction des études, à la délivrance des diplômes.

L'évaluation formative, quant à elle, est «un processus d'évaluation continue ayant pour objet d'assurer la progression de chaque individu dans une démarche d'apprentissage avec l'intention de modifier le rythme de cette progression ou la démarche elle-même, pour apporter (s'il y a lieu) des améliorations ou des correctifs appropriés ²».

Autrement dit, l'évaluation formative est composée de tous les moyens que vous prenez pour aider l'élève à apprendre pendant qu'il apprend. Ces moyens peuvent être différents d'un élève à l'autre puisque ces derniers sont différents.

-
1. MEQ (Ministère de l'Éducation du Québec), **Éléments de docimologie, lexique**, fascicule 2, 1985, p. 10, 11.
 2. Gérard Scallon, **Définition, exigences et caractéristiques de l'évaluation formative**, notes de cours.

ÉVALUATION SOMMATIVE ET ÉVALUATION FORMATIVE

	<u>Évaluation sommative</u>	<u>Évaluation formative</u>
<i>Intention de l'enseignant</i>	<i>Sanctionner des apprentissages.</i>	<i>Favoriser la progression de l'apprentissage.</i>
<i>Objet de l'évaluation</i>	<i>Les objectifs terminaux du programme d'études (apprentissage terminé).</i>	<i>Des préalables, la démarche de l'élève, des apprentissages inachevés ou modifiables, des stratégies.</i>
<i>Scénario</i>	<i>Rigide, semblable pour tous les élèves soumis à des conditions identiques.</i>	<i>Souple, parfois différent d'un élève à l'autre. La participation de l'élève peut être sollicitée.</i>
<i>Moment</i>	<i>À la fin d'une période d'enseignement et d'apprentissage.</i>	<i>Pendant la période d'enseignement et d'apprentissage.</i>

ÉVALUATION FORMATIVE ET ÉVALUATION SOMMATIVE

Pour chacun des énoncés suivants, indiquez (EF) s'il appartient à l'évaluation formative ou indiquez (ES) s'il appartient à l'évaluation sommative.

1. Un élève éprouve certaines difficultés en sciences; son professeur le rencontre pour sonder son degré de motivation à l'étude. ()
2. Une épreuve de sciences est administrée à la fin d'une étape; les scores sont convertis en rangs pour être consignés dans un bulletin. ()
3. Les résultats de l'épreuve de juin 1991 sont utilisés en septembre 1991 pour préparer des activités de révision.
(Répondre en fonction des élèves qui ont subi l'épreuve.) ()
4. Un enseignant interroge oralement ses élèves afin de décider s'il reviendra sur certaines notions. ()
5. Une enseignante administre une épreuve dont la cible sera des apprentissages inachevés. ()
6. Dans le but de sanctionner des apprentissages, une épreuve portant sur l'ensemble des objectifs du programme de sciences est administrée à la fin de l'année. ()
7. Dans le but de favoriser la progression de l'apprentissage, une épreuve portant sur un seul objectif est administrée pendant que les élèves réalisent cet apprentissage. ()
8. Les élèves subissent, à chaque semaine, une épreuve de rendement; les scores de ces épreuves sont additionnés de façon à produire pour chaque élève un bilan qui servira d'évaluation à la fin du cours. ()

4. L'INTERPRÉTATION NORMATIVE ET L'INTERPRÉTATION CRITÉRIELLE

Interpréter un score de façon critérielle consiste essentiellement à comparer ce score à un critère fixé à l'avance et pris comme seuil de réussite.

Interpréter un score de façon normative consiste essentiellement à comparer ce score à une norme (moyenne de groupe).

En évaluation sommative, on peut interpréter les résultats de façon normative ou critérielle.

En évaluation formative, on interprète les résultats de façon critérielle.

INTERPRÉTATION NORMATIVE OU INTERPRÉTATION CRITÉRIELLE

Pour chacun des énoncés suivants, indiquez (IN) si les résultats sont interprétés de façon normative ou indiquez (IC) si les résultats sont interprétés de façon critérielle.

1. À une compétition de ski, on remet une médaille aux trois personnes qui ont été les premières à franchir la ligne d'arrivée. ()
2. À la compétition des BONS VIVANTS, on remet un prix à tous les skieurs qui ont pris moins de cinq minutes à franchir la ligne d'arrivée et ce, sans tomber. ()
3. À une épreuve de sciences, seuls les élèves qui auront réussi huit problèmes sur dix pourront aborder le chapitre suivant. ()
4. À l'épreuve de juin en sciences, les dix élèves les plus forts obtiendront la lettre A. ()
5. Comparer les élèves entre eux pour choisir le plus grand. ()
6. D'après un règlement du ministère de l'Éducation, seuls les enfants ayant atteint l'âge prescrit au 30 septembre seront autorisés à commencer leurs études primaires. ()
7. Pour interpréter le résultat d'un élève à une épreuve sommative, son score est transformé en rang stanine. ()



*ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE*

ACTIVITÉ 2

Les qualités des instruments de mesure

GUIDE DE L'ANIMATEUR

ACTIVITÉ 2

Les qualités des instruments de mesure

1. La validité
2. La fidélité
3. La congruence
4. La pertinence
5. La difficulté de l'item ou de l'épreuve
6. Pourquoi bâtir une épreuve en fonction d'une définition du domaine?

1. LA VALIDITÉ

La qualité la plus importante.

L'ensemble des items de votre épreuve représente-t-il bien l'ensemble des objectifs que vous voulez mesurer?

Si oui, votre épreuve est valide.

Les moyens de s'assurer de la validité d'une épreuve :

- déterminer clairement les objectifs que l'on veut mesurer;*
- se faire un tableau de spécifications (définition du domaine);*
- s'assurer de la congruence entre l'objectif dont on veut mesurer l'atteinte et le contenu des items;*
- voir à ce que les items soient bien construits.*

ACÉTATE 1

LA VALIDITÉ

Premier indice afin de reconnaître la qualité de nos instruments de mesure.

Qualité la plus importante de l'épreuve.

L'ensemble des items de votre épreuve représente-t-il bien l'ensemble des objectifs que vous voulez mesurer?

Si oui, votre épreuve est valide.

Les moyens de s'assurer de la validité d'une épreuve :

☐

déterminer clairement les objectifs terminaux que l'on veut mesurer;

☐

se faire un tableau de spécifications;

☐

s'assurer de la congruence entre l'objectif terminal dont on veut mesurer l'atteinte et le contenu des items de l'épreuve;

☐

voir à ce que les items soient bien construits.

2. LA FIDÉLITÉ

Les résultats de l'épreuve sont-ils stables?

«Est-ce que le résultat de Justine ressemble à Justine?»

Erreur de mesure (E) faible.

Les moyens de s'assurer de la fidélité d'une épreuve :

- porter attention aux erreurs inhérentes à l'épreuve:
 - longueur de l'épreuve,
 - vocabulaire utilisé,
 - ambiguïté de l'épreuve,
 - temps limité,
 - correction (deux correcteurs = même note);
- porter attention aux erreurs dues à la passation de l'épreuve:
 - conditions identiques pour tous, etc.;
- porter attention aux erreurs provenant de l'élève lui-même (les plus difficiles à contrôler).

ACÉTATE 2

LA FIDÉLITÉ

Deuxième indice afin de reconnaître la qualité de nos instruments de mesure.

Les résultats de l'épreuve sont-ils stables?

«Est-ce que le résultat de Justine ressemble à Justine?»

Erreur de mesure (E) faible.

Les moyens de s'assurer de la fidélité d'une épreuve :

☐

porter attention aux erreurs inhérentes à l'épreuve;

- longueur de l'épreuve,
- vocabulaire utilisé,
- ambiguïté de l'épreuve,
- temps limité,
- correction (deux correcteurs = même note);

☐

porter attention aux erreurs dues à la passation de l'épreuve;

- conditions identiques pour tous, etc.;

☐

porter attention aux erreurs provenant de l'élève lui-même (les plus difficiles à contrôler).

3. LA CONGRUENCE

ACÉTATE 3

LA CONGRUENCE

La congruence est la relation étroite qui doit exister entre un item et un objectif.

Pour qu'un item soit considéré comme congruent avec un objectif, il doit respecter trois critères :

- 1. Une relation étroite doit exister entre le verbe d'action présent dans l'objectif terminal et la tâche de l'item à accomplir.*
- 2. Les conditions définies dans l'objectif terminal et celles présentées dans l'item à accomplir doivent être les mêmes.*
- 3. Le contenu que l'on retrouve dans l'item doit être ou bien un échantillon du contenu décrit dans l'objectif terminal, ou bien le contenu en entier de cet objectif.*

Exercice :

Présenter sur acétate un objectif suivi d'un item.

Demander aux participants de juger de la congruence de l'item par rapport à l'objectif en tenant compte des trois critères énoncés à la page précédente.

- 1. Demander d'indiquer le verbe, le contenu et la condition (si présente) de l'objectif.*
- 2. Demander si le verbe de l'objectif est congruent avec la tâche de l'item à accomplir.*
- 3. Demander si le contenu (en partie ou en totalité) est congruent avec le contenu décrit dans l'item.*
- 4. Demander si la condition (si présente) de l'objectif est décrite dans l'item.*
- 5. Si l'un ou l'autre des critères n'est pas respecté, demander aux participants d'améliorer l'item afin de le rendre parfaitement congruent avec l'objectif.*

ACÉTATE 4 A)

OBJECTIF TERMINAL

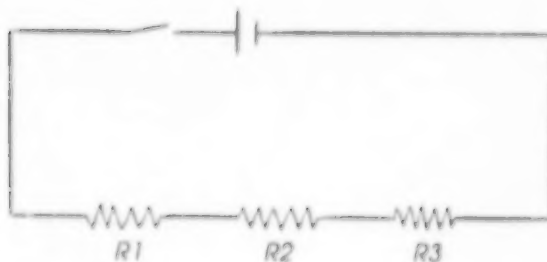
Analyser les variables caractéristiques de diverses associations d'éléments de circuits électriques utilisés au moment des travaux scientifiques.

OBJECTIF INTERMÉDIAIRE

Déterminer la résistance équivalente de circuits en série et de circuits en parallèle, à température constante.

ITEM

D'après le circuit suivant, quelle est la relation mathématique qui permet de calculer la résistance équivalente?



4. LA PERTINENCE

Caractère de ce qui est plus ou moins approprié, qui s'inscrit dans la ligne de l'objectif poursuivi ou de la dimension.

5. LA DIFFICULTÉ DE L'ITEM ET DE L'ÉPREUVE

La qualité d'une épreuve ou d'un item ne dépend pas de sa difficulté.

Bien souvent le facteur difficulté dépend de l'objectif même. L'important, c'est que l'item soit congruent avec l'objectif dont il mesure l'atteinte. La difficulté de la tâche à accomplir va dépendre de la difficulté de l'objectif.

Tout ceci est d'autant plus vrai lorsque les résultats de l'épreuve servent à prendre des décisions relativement à la sanction des études.

6. **POURQUOI BÂTIR UNE ÉPREUVE EN FONCTION D'UNE DÉFINITION
DU DOMAINE?**

On a vu l'importance de la VALIDITÉ d'une épreuve quand on se préoccupe de bâtir des instruments de qualité.

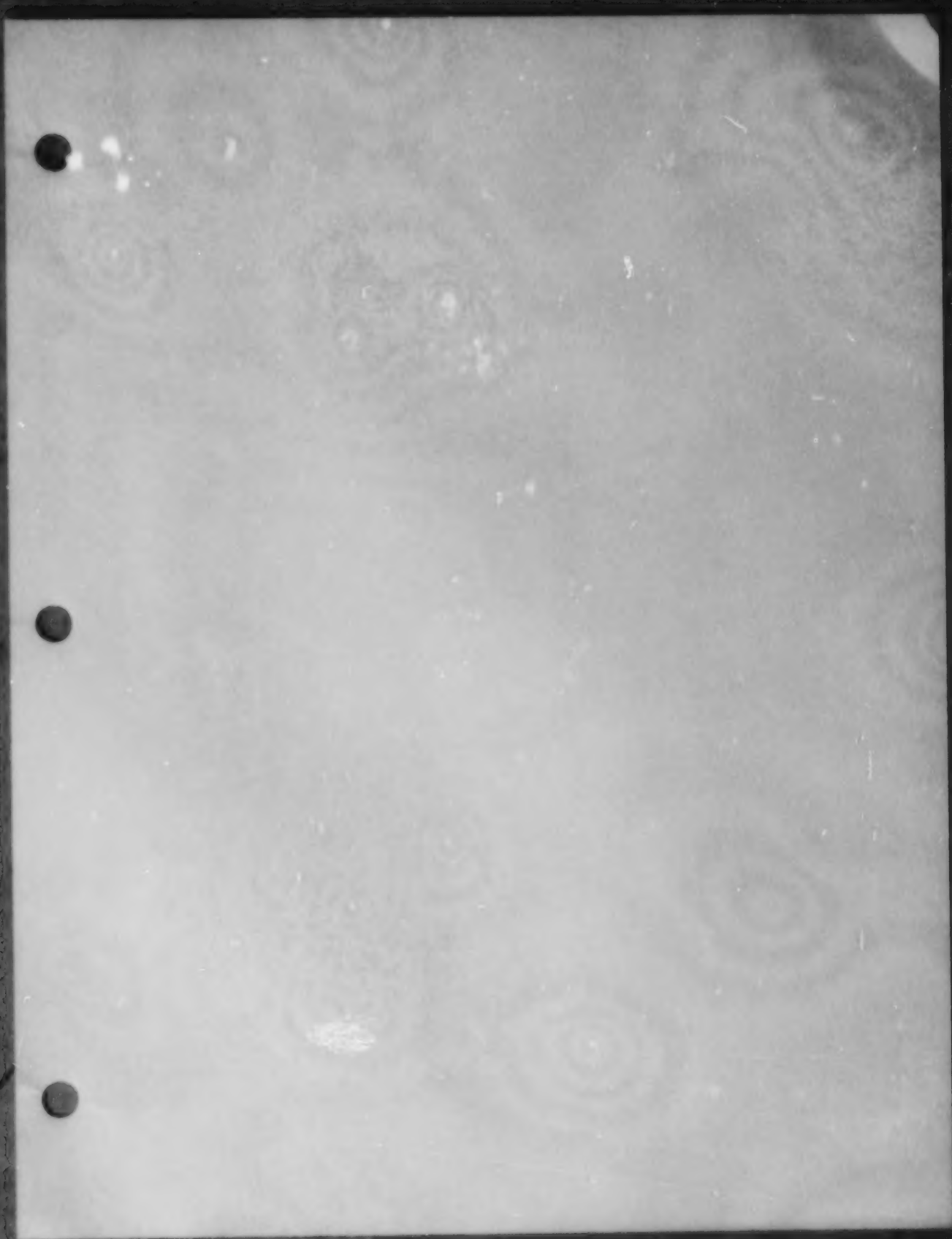
La définition du domaine a donc été décrite comme un moyen efficace d'obtenir une synthèse des éléments essentiels d'un programme d'études pour des fins d'évaluation sommative.

Pour ce, il faut se rappeler que la définition du domaine vise essentiellement à assurer la congruence entre un programme d'études et tout instrument de mesure devant servir à une évaluation sommative.

Une synthèse :

ACÉTATE 5

**ÉPREUVE À L'ÉCHELLE
PROVINCIALE***(de type papier-crayon)*↓
DOMAINE↓
DIMENSION↓
ITEMS**PROGRAMME D'ÉTUDES***(sans les contraintes
occasionnées par une épreuve
à l'échelle provinciale)*↓
**ORIENTATION ET PRINCIPES
DIRECTEURS**↓
**UN SOUS-ENSEMBLE D'OBJECTIFS
TERMINAUX**↓
PERTINENCE↓
CHOISIR L'OBJECTIF↓
CONGRUENCE



ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 2
Les qualités des instruments de mesure
GUIDE DU PARTICIPANT

ACTIVITÉ 2

Les qualités des instruments de mesure

1. *La validité*
2. *La fidélité*
3. *La congruence*
4. *La pertinence*
5. *La difficulté de l'item ou de l'épreuve*
6. *Pourquoi bâtir une épreuve en fonction d'une définition du domaine?*

1. LA VALIDITÉ

Premier indice afin de reconnaître la qualité de nos instruments de mesure.

Qualité la plus importante de l'épreuve.

L'ensemble des items de votre épreuve représente-t-il bien l'ensemble des objectifs que vous voulez mesurer?

Si oui, votre épreuve est valide.

Les moyens de s'assurer de la validité d'une épreuve :

☐

déterminer clairement les objectifs terminaux que l'on veut mesurer;

☐

se faire un tableau de spécifications;

☐

s'assurer de la congruence entre l'objectif terminal dont on veut mesurer l'atteinte et le contenu des items de l'épreuve;

☐

voir à ce que les items soient bien construits.

2. LA FIDÉLITÉ

Deuxième indice afin de reconnaître la qualité de nos instruments de mesure.

Les résultats de l'épreuve sont-ils stables?

«Est-ce que le résultat de Justine ressemble à Justine?»

Erreur de mesure (E) faible.

Les moyens de s'assurer de la fidélité d'une épreuve :

☐

porter attention aux erreurs inhérentes à l'épreuve:

- longueur de l'épreuve,
- vocabulaire utilisé,
- ambiguïté de l'épreuve,
- temps limité,
- correction (deux correcteurs = même note);

☐

porter attention aux erreurs dues à la passation de l'épreuve:

- conditions identiques pour tous, etc.;

☐

porter attention aux erreurs provenant de l'élève lui-même (les plus difficiles à contrôler).

3. LA CONGRUENCE

La congruence est la relation étroite qui doit exister entre un item et un objectif.

Pour qu'un item soit considéré comme congruent avec un objectif, il doit respecter trois critères :

1. *Une relation étroite doit exister entre le verbe d'action présent dans l'objectif terminal et la tâche de l'item à accomplir.*
2. *Les conditions définies dans l'objectif terminal et celles présentées dans l'item à accomplir doivent être les mêmes.*
3. *Le contenu que l'on retrouve dans l'item doit être ou bien un échantillon du contenu décrit dans l'objectif terminal, ou bien le contenu en entier de cet objectif.*

Exercice sur la congruence

4. LA PERTINENCE

Caractère de ce qui est plus ou moins approprié, qui s'inscrit dans la ligne de l'objectif poursuivi ou de la dimension.

5. LA DIFFICULTÉ DE L'ITEM ET DE L'ÉPREUVE

La qualité d'une épreuve ou d'un item ne dépend pas de sa difficulté.

Bien souvent le facteur difficulté dépend de l'objectif même. L'important, c'est que l'item soit congruent avec l'objectif dont il mesure l'atteinte. La difficulté de la tâche à accomplir va dépendre de la difficulté de l'objectif.

Tout ceci est d'autant plus vrai lorsque les résultats de l'épreuve servent à prendre des décisions relatives à la sanction des études.

6. **POURQUOI BÂTIR UNE ÉPREUVE EN FONCTION D'UNE DÉFINITION
DU DOMAINE?**

On a vu l'importance de la VALIDITÉ d'une épreuve quand on se préoccupe de bâtir des instruments de qualité.

La définition du domaine a donc été décrite comme un moyen efficace d'obtenir une synthèse des éléments essentiels d'un programme d'études pour des fins d'évaluation sommative.

Pour ce, il faut se rappeler que la définition du domaine vise essentiellement à assurer la congruence entre un programme d'études et tout instrument de mesure devant servir à une évaluation sommative.

UNE SYNTHÈSE

ÉPREUVE À L'ÉCHELLE PROVINCIALE

(de type papier-crayon)

↓
DOMAINE

↓
DIMENSION

↓
ITEMS

PROGRAMME D'ÉTUDES

(sans les contraintes
occasionnées par une épreuve
à l'échelle provinciale)

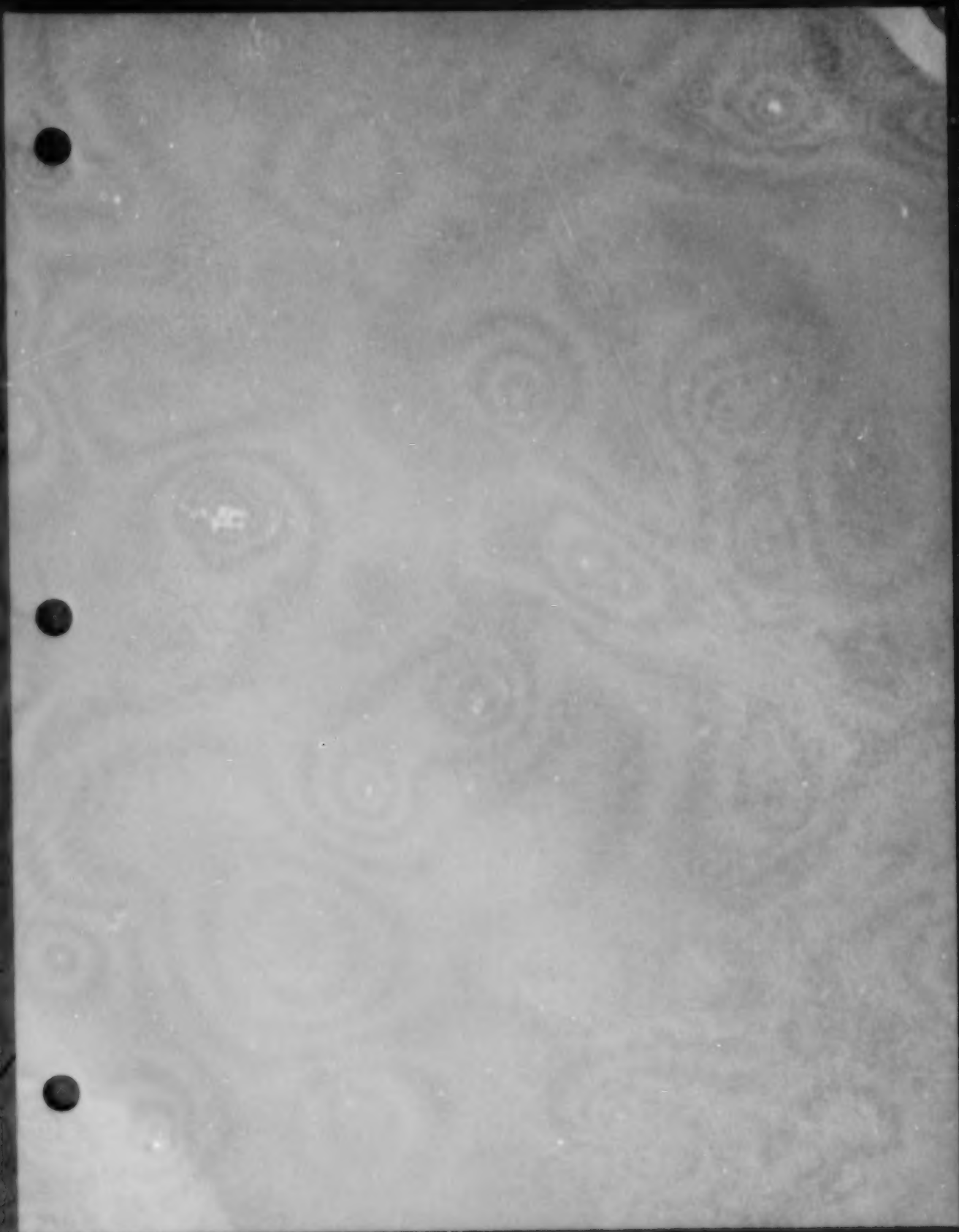
↓
ORIENTATION ET PRINCIPES
DIRECTEURS

↓
UN SOUS-ENSEMBLE D'OBJECTIFS
TERMINAUX

↓
PERTINENCE

↓
CHOISIR L'OBJECTIF

↓
CONGRUENCE



ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 3

Les définitions du domaine en sciences
GUIDE DE L'ANIMATEUR

ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 3

Les définitions du domaine en sciences

GUIDE DE L'ANIMATEUR

ACTIVITÉ 3

Les définitions du domaine en sciences

1. *Les thèmes*
2. *Les habiletés et leur limite*
3. *Les dimensions*
4. *Les items et leur dimension*

1. LES THÈMES

*Les contenus des programmes d'études sont synthétisés
et regroupés à l'aide de différents thèmes.*

2. LES HABILETÉS ET LEUR LIMITE

Avant que les responsables des programmes d'études définissent les habiletés propres à chaque programme, à l'intérieur des définitions du domaine, on avait recours à différentes taxinomies. Celle de Bloom était sans doute la plus connue et la plus utilisée. Pour certains programmes, la taxinomie de Bloom convenait très bien, mais pour d'autres, elle convenait moins bien. C'est pour cette raison qu'il y a eu une volonté de recourir à des habiletés propres à chacun des programmes d'études. Ces habiletés sont maintenant décrites dans les définitions du domaine de chaque programme.

Une habileté, en ce qui a trait à la tâche à accomplir, est le scénario le plus probable pour la majorité des élèves. On pourrait parler d'une habileté comme d'une dominante.

En sciences physiques, on a défini quatre habiletés. Trois de ses quatre habiletés ne sont pas exclusives en soi.

Cette limite aura des répercussions sur les dimensions.

ACÉTATES 1 et 2

CARACTÉRISER :

Préciser des faits, des phénomènes de l'environnement, une procédure et des techniques, dans le contexte des cheminements d'apprentissage suggérés par le programme.

RELIER :

Établir des rapports entre des faits et des phénomènes, dans le contexte des cheminements d'apprentissage suggérés par le programme.

RÉSoudre DES PROBLÈMES EN CLASSE :

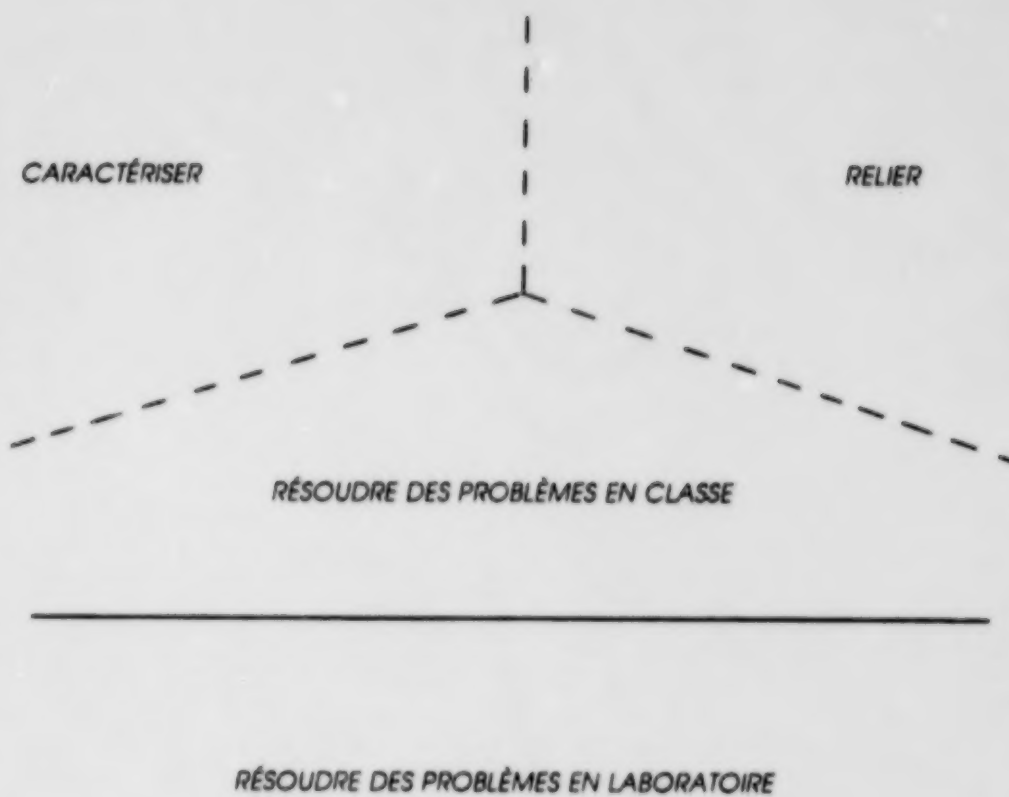
Organiser les composantes de la démarche de résolution de problèmes ou de la méthode scientifique et de l'apprentissage pour solutionner des problèmes en classe.

RÉSoudre DES PROBLÈMES EN LABORATOIRE :

Organiser les composantes de la méthode (démarche) scientifique et de l'apprentissage pour solutionner des problèmes en laboratoire.

Note : *S'en tenir aux définitions des habiletés à l'intérieur des définitions du domaine.*

LES HABILETÉS ET LEUR LIMITE



3. LES DIMENSIONS

Les dimensions sont obtenues par le croisement d'une habileté et d'un thème.

Ces dimensions ne sont pas exclusives en soi.

ACÉTATE 3-3

Le tableau de synthèse

LES SCIENCES PHYSIQUES 416

MODULE	STRUCTURE ET MATIÈRE (33 %)	PHÉNOMÈNES ÉLECTRIQUES (40 %)	PHÉNOMÈNES IONIQUES (27 %)
HABILETÉ			
CARACTÉRISER (38 %)			
RELIER (40 %)			
RÉSoudre DES PROBLÈMES EN CLASSE (22 %)			
RÉSoudre DES PROBLÈMES EN LABORATOIRE			

4. LES ITEMS ET LEUR DIMENSION

Exercice :

À l'aide de l'acétate 3-3 représentant le tableau de synthèse et d'une série d'items, faire associer, par les participants, les items à leur dimension.

En assemblée plénière, préciser les habiletés et les dimensions, amener la limite des habiletés et des dimensions.

Pour l'exercice, on a constitué une banque d'items qui vous sont présentés dans les pages suivantes.

Il est à noter que ces items ne sont pas présentés comme modèles.

Vous êtes invités à ajouter vos propres items.

LE PROBLÈME

Situation

Dernièrement, on vous a embauché comme analyste-chimiste dans une industrie où un accident s'est produit.

Pour des raisons que l'on ignore, un gaz s'est formé après le contact d'un liquide et d'un solide inconnus. Plusieurs personnes ont été atteintes par le gaz émis.

Les dirigeants de l'industrie veulent connaître avec **précision** les causes de cet accident.

Questions

Quels sont les produits (solide et liquide inconnus) qui ont été mis en contact?

Quelles sont les trois propriétés du solide et celles du liquide mis en présence l'un de l'autre (utilisez, au besoin, l'extrait du vade-mecum en annexe)?

Quelle est la nature du gaz produit, son caractère particulier?

Consignes

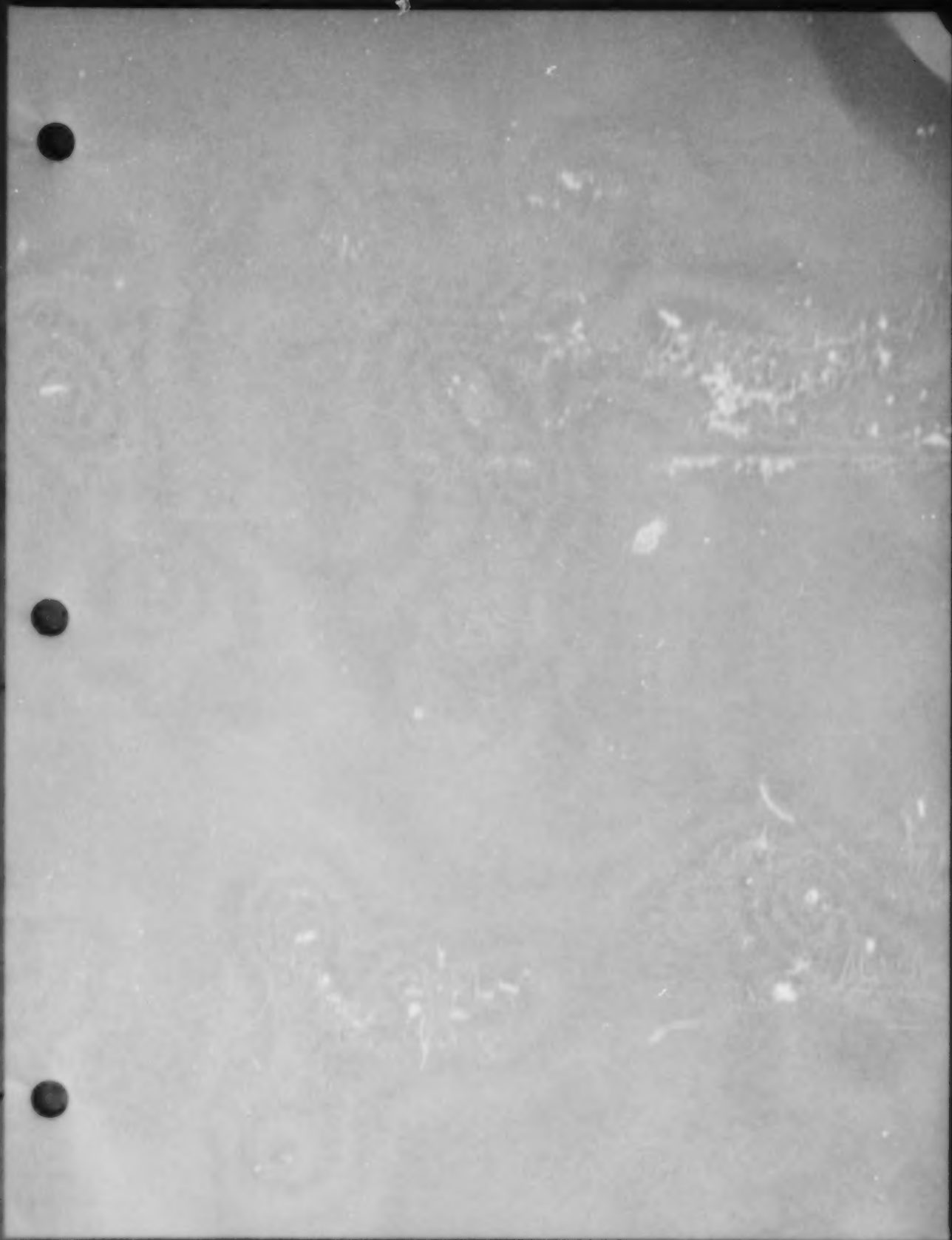
Des échantillons du liquide et du solide vous seront remis.

Vous devez justifier vos conclusions dans le cahier de réponses qu'on vous remettra, en retraçant les étapes de la méthode expérimentale que vous aurez appliquée.

Le tableau suivant indique le numéro des questions qui appartiennent à chacune des dimensions du projet de définition du domaine.

DISTRIBUTION DES QUESTIONS

MODULE HABILETÉ	STRUCTURE ET MATIÈRE	PHÉNOMÈNES ÉLECTRIQUES	PHÉNOMÈNES IONIQUES
CARACTÉRISER	1, 2, 3	5, 6, 7, 8, 9	10, 11
METTRE EN RELATION	4, 14, 15	16, 17, 21	18, 19
APPLIQUER	12, 20	22, 23, 24	13, 25



1. Dans le tableau de classification périodique, on trouve les éléments suivants :

18 Ar
Argon

19 K
Potassium

25 Mn
Manganèse

35 Br
Brome

Que nous révèle la position de ces éléments dans le tableau de classification périodique?

(A) 18 Ar
Argon est un gaz inerte.

19 K
Potassium est un alcalin.

25 Mn
Manganèse est un métal.

35 Br
Brome est un non-métal.

(B) 18 Ar
Argon est un non-métal.

19 K
Potassium est un métal.

25 Mn
Manganèse est un métal.

35 Br
Brome est un alcalino-terreux.

(C) 18 Ar
Argon est un gaz inerte.

19 K
Potassium est un métal.

25 Mn
Manganèse est un halogène.

35 Br
Brome est un non-métal.

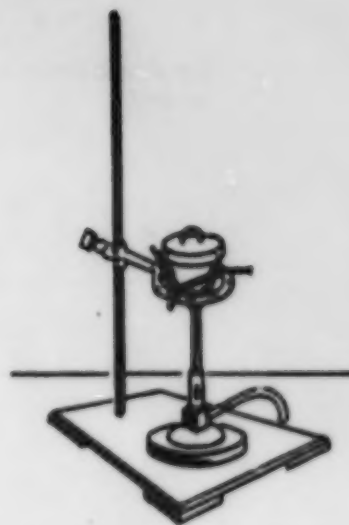
(D) 18 Ar
Argon est un halogène.

19 K
Potassium est un alcalin.

25 Mn
Manganèse est un métal.

35 Br
Brome est un non-métal.

2. Au laboratoire, on a chauffé de petits morceaux de calcaire dans un creuset à 1000°C pendant 15 minutes et on les a laissés refroidir. Avant et après le chauffage, on a procédé aux tests suivants : mesure de la masse, observation de la couleur des morceaux, effet de l'eau et effet de l'acide acétique (vinaigre). On a obtenu les résultats suivants :



Test	Avant le chauffage	Après le chauffage
Masse	10,0 g	5,7
Couleur	gris foncé	blanc
Effet de l'eau	aucun	effervescence
Effet du vinaigre	effervescence	aucun

Que peut-on déduire de ces résultats?

- (A) Il y a eu un changement physique parce que la nature de la substance n'a pas été modifiée.
- (B) Il y a eu un changement chimique parce que la nature de la substance a été modifiée.
- (C) Il y a eu un changement physique parce que la nature de la substance a été modifiée.
- (D) Il y a eu un changement chimique parce que la nature de la substance n'a pas été modifiée.

3. Pauline a chauffé un mélange de poudre de soufre et de plomb dans une éprouvette fermée par une membrane de caoutchouc. Après la réaction, elle observe du soufre sur les parois de l'éprouvette et du sulfure de plomb au fond.

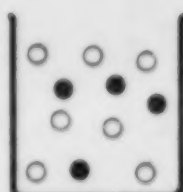
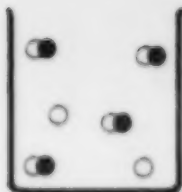
Quel modèle représente cette transformation chimique?

Note. On représente le soufre par $\circ$ et le plomb par $\bullet$.

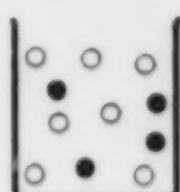
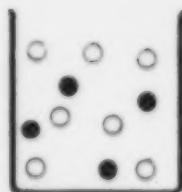
Avant réaction

Après réaction

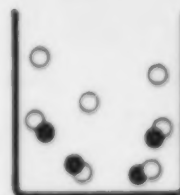
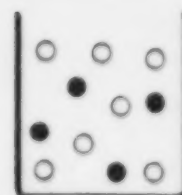
(A)



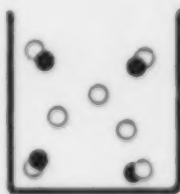
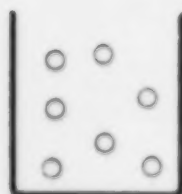
(B)



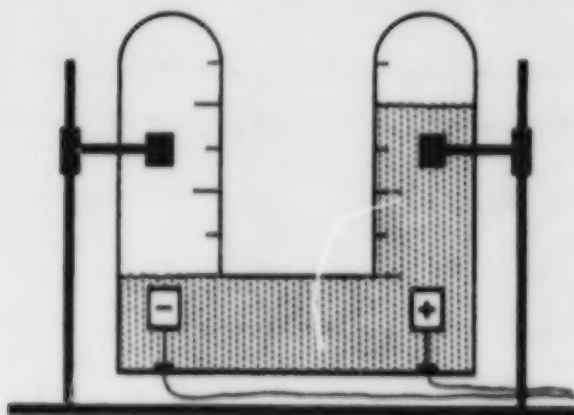
(C)



(D)



4. Après l'électrolyse d'une substance inconnue, on mesure 30 ml d'un gaz x et 10 ml d'un gaz y.



Lequel des modèles suivants représente la molécule de la substance inconnue?

(A)



(C)



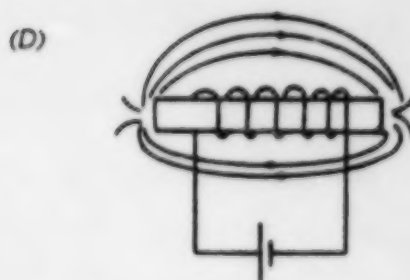
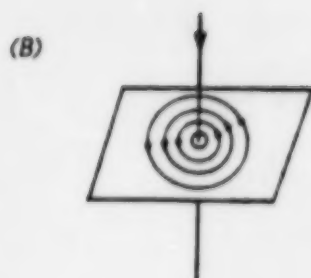
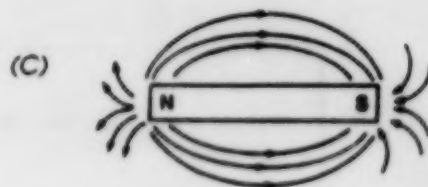
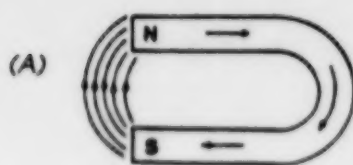
(B)



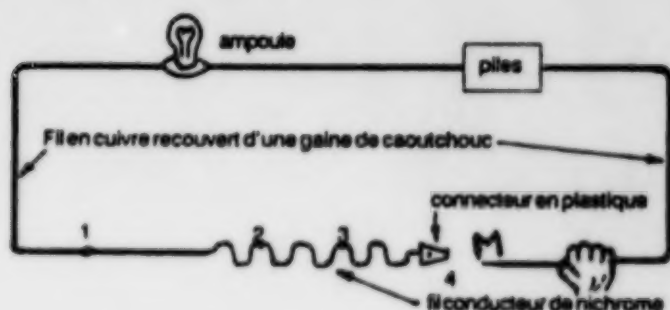
(D)



5. Lequel des schémas ci-dessous est une représentation correcte du champ magnétique?



6. Pour varier l'intensité de la lumière d'une ampoule, Stéphane a fait le montage électrique suivant :



L'extrémité M peut toucher aux points de contact 1, 2, 3 ou 4.

Quel point de contact donne la lumière la plus intense?

- (A) Le point 1 situé sur la gaine de caoutchouc.
- (B) Le point 2 situé sur le fil de nichrome.
- (C) Le point 3 situé sur le fil de nichrome.
- (D) Le point 4 situé sur le connecteur en plastique.

7. Liz et François étudient le circuit électrique illustré ci-dessous.



Ils mesurent la tension aux bornes de chacun des résistors. Ils en concluent que la tension aux bornes de la source est égale à la somme des tensions aux bornes de chacun des résistors du circuit.

Êtes-vous d'accord avec cette affirmation et pourquoi?

- (A) Oui, parce que c'est un circuit en série.
- (B) Oui, parce que c'est un circuit en parallèle.
- (C) Non, parce que c'est un circuit en série.
- (D) Non, parce que c'est un circuit en parallèle.

8. Pascal fait fonctionner quatre appareils électriques pendant une heure. Les plaques signalétiques ci-dessous montrent les caractéristiques de chaque appareil.

Quel appareil a consommé le plus d'énergie électrique?

(A) Micro-ordinateur

120 V
216 W
1,8 A

(C) Radio

120 V
36 W
0,3 A

(B) Jeu vidéo

120 V
18 W
0,15 A

(D) Téléviseur

120 V
240 W
2,0 A

9. À Tracy, on fait brûler de l'huile pour produire de l'électricité. Certaines personnes prétendent que ce procédé n'a pas d'effet sur l'environnement.

Êtes-vous d'accord avec cette affirmation et pourquoi?

- (A) Non, parce que la combustion dégage des gaz qui acidifient les pluies et du gaz carbonique.
- (B) Oui, parce que l'huile brûle sans produire de fumée visible.
- (C) Non, parce que la combustion produit un déchet radioactif.
- (D) Oui, parce que la chaleur dégagée par la fumée réchauffe l'atmosphère avoisinante.

10. Un produit nettoyant contient de l'ammoniaque. On veut déterminer expérimentalement si l'ammoniaque est un acide, une base ou un sel.

Au cours de cette expérience, on doit vérifier certaines propriétés parmi les suivantes :

- a) effet sur le métal;
- b) effet sur le papier de chlorure de cobalt;
- c) effet sur le papier tournesol;
- d) trouble de l'eau de chaux;
- e) effet sur un acide ou une base;
- f) activation (raviver) des tisons;
- g) explosion en présence d'une flamme.

Lesquelles de ces propriétés doit-on vérifier?

- | | |
|---------------|---------------|
| (A) a, c et e | (C) b, d et g |
| (B) b, c et d | (D) e, f et g |

11. Kim et Sébastien mesurent le pH de différentes solutions à l'aide d'un indicateur universel.

Ils ont noté les résultats suivants :

Solution	pH
Eau salée	8
Eau gazeuse	3
Nettoyeur liquide	11
Lave-vitre	9
Antiacide	10
Jus de raisin vert	3
Jus de pomme de terre	6
Vinaigre	3

Quelles sont les solutions acides?

- (A) Nettoyeur liquide, jus de raisin vert, antiacide et eau salée.
- (B) Eau gazeuse, jus de raisin vert, jus de pomme de terre et vinaigre.
- (C) Nettoyeur liquide, lave-vitre, eau gazeuse et vinaigre.
- (D) Jus de pomme de terre, lave-vitre, antiacide et eau salée.

12. En laboratoire, on a fait réagir un élément inconnu X avec le fluor.

On obtient les composés suivants : XF

D'après les électrons de valence du fluor, dans quelle famille classez-vous l'élément X?

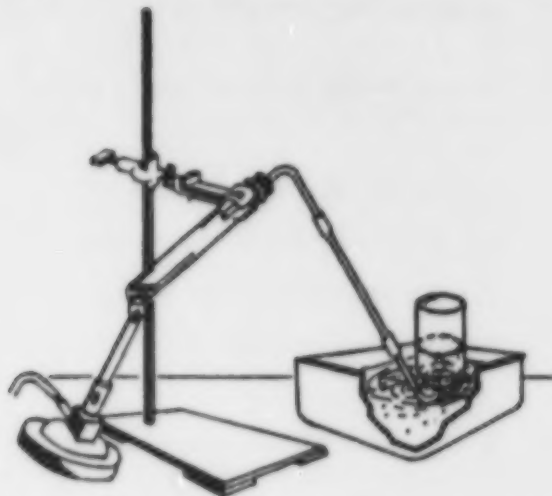
- (A) Dans la famille des halogènes.
- (B) Dans la famille des gaz rares.
- (C) Dans la famille des alcalins.
- (D) Dans la famille des alcalino-terreux.

13. Au laboratoire, on a neutralisé une solution d'hydroxyde de potassium (KOH) avec de l'acide sulfurique (H_2SO_4).

Quelle équation équilibrée représente correctement la réaction de neutralisation?

- (A) $\text{KOH} + \text{H}_2\text{SO}_4 \rightarrow \text{H}_2\text{O} + \text{KSO}_4$
- (B) $\text{KOH} + \text{H}_2\text{SO}_4 \rightarrow \text{H}_2\text{O} + \text{K}_2\text{SO}_4$
- (C) $\text{KOH} + \text{H}_2\text{SO}_4 \rightarrow \text{OH}\text{SO}_4 + \text{KH}_2$
- (D) $2 \text{KOH} + \text{H}_2\text{SO}_4 \rightarrow 2 \text{H}_2\text{O} + \text{K}_2\text{SO}_4$

14. Claude chauffe dans une éprouvette une substance solide blanche. Elle observe un dégagement de gaz et il reste un solide gris dans l'éprouvette.

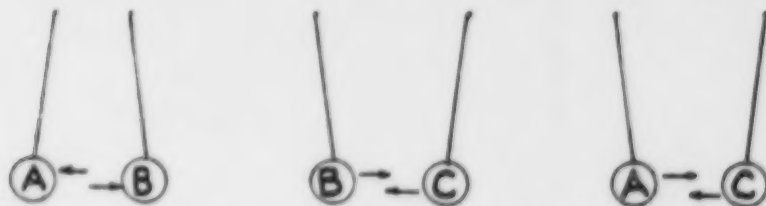


La substance de départ est-elle un composé ou un élément? Justifiez votre réponse.

15. On a suspendu deux boules A et B chargées.

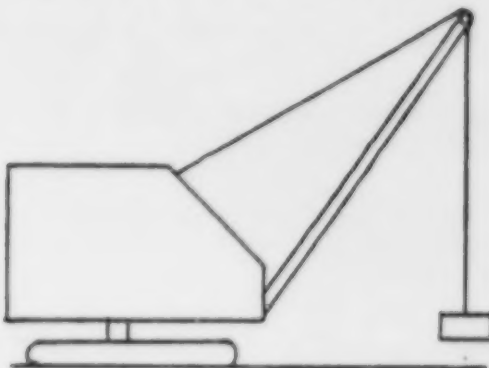
Quand elles sont près l'une de l'autre, on observe une répulsion. On approche une boule C chargée, de la boule B : on observe une attraction.

Lorsqu'on place la boule C près de la boule A, il y a attraction.



Que peut-on conclure de cette expérience à propos des charges des boules A, B et C?

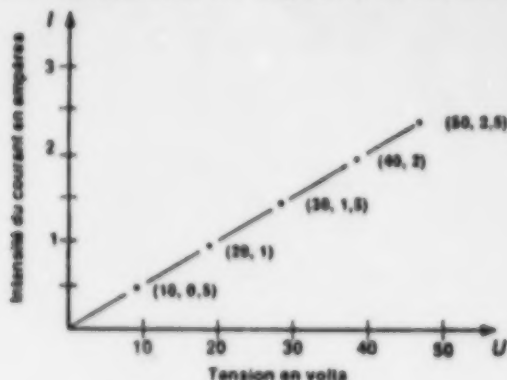
16. Dans une aciérie, on utilise une grue mécanique munie d'un électro-aimant puissant suspendu au bout d'un câble pour charger des camions de pièces de ferraille de toutes dimensions et de toutes formes.



Pourquoi utilise-t-on un électro-aimant plutôt qu'un aimant naturel?

17. Karl voudrait remplacer un résistor de sa radio, mais il en ignore la conductance.

Il mesure alors différentes intensités de courant lorsque la tension varie. Les valeurs sont représentées sur le graphique suivant :



Quelle est la conductance du résistor?

18. Parmi les composés suivants, on trouve des acides, des bases et des sels :



Classez chaque composé, d'après la formule moléculaire, dans la colonne appropriée.

19. Dans la neutralisation de l'acide chlorhydrique HCl par de l'hydroxyde de sodium $\text{Mg}(\text{OH})_2$, il se forme du chlorure de magnésium, MgCl_2 , et de l'eau, H_2O .

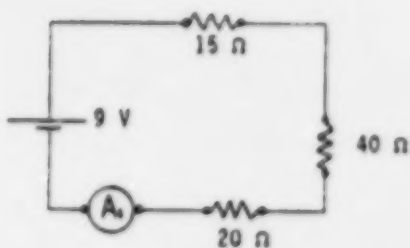
Quelle équation équilibrée correspond à la réaction de neutralisation?

20. On a retiré les étiquettes de trois bonbonnes contenant respectivement de l'oxygène, de l'hydrogène et du dioxyde de carbone.

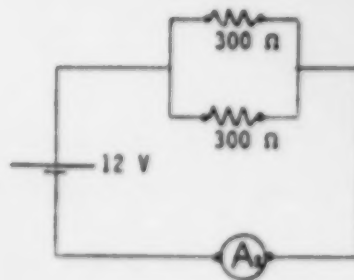
Élaborez un protocole qui vous permettra de remettre chaque étiquette sur la bonbonne correspondante.

Indiquez toutes les étapes de votre démarche.

21. Deux circuits électriques comprennent différents résistors montés différemment.



Circuit I

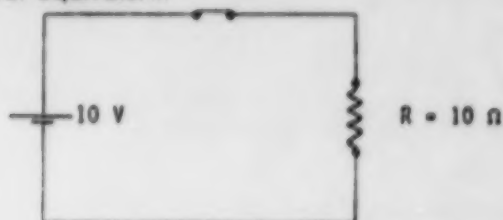


Circuit II

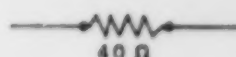
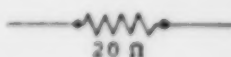
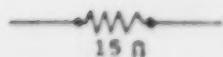
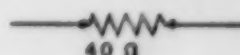
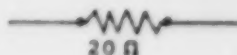
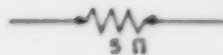
Quel ampèremètre indique la plus grande intensité de courant?

Indiquez toutes les étapes de votre démarche.

22. Le résistor R du circuit électrique illustré ci-dessous doit être remplacé par un résistor équivalent.



Vous disposez seulement des six résistors de remplacement suivants :

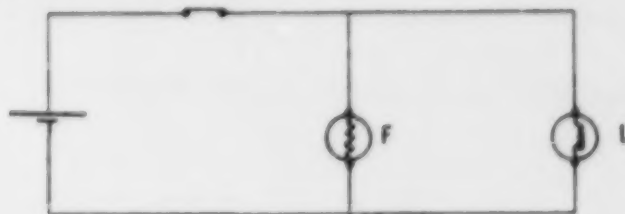


Quel montage vous permet d'avoir un résistor équivalent au résistor R ?

Indiquez toutes les étapes de votre démarche.

23. François allume une ampoule (L) de 60 watts et repasse ses vêtements avec un fer à repasser (F) de 600 watts. La durée du repassage est de 1 h 30.

Hydro-Québec facture 0,048 \$/kWh.

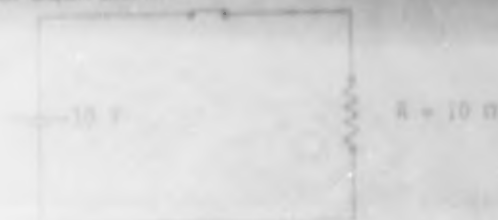


Combien a coûté l'usage de ces deux appareils?

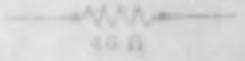
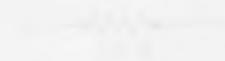
Indiquez toutes les étapes de votre démarche.

22.

Le résistor R du circuit électrique illustré ci-dessous doit être remplacé par un résistor équivalent.



Vous disposez actuellement des six résistors de remplacement suivants :



Quel montage vous permet d'avoir un résistor équivalent au résistor R ?

Indiquez toutes les étapes de votre démarche.

23.

On peut déformer une ampoule (L) de 60 watts et repasser ses filaments avec un fer à repasser (F) de 600 watts. La durée du chauffage est de 1 h 30.

Donc Québec facture 0,048 \$/kWh.



Combien a coûté l'usage de ces deux appareils?

Indiquez toutes les étapes de votre démarche.

24. Un élément chauffant baigne dans un becher contenant 1000 grammes d'eau. On applique une certaine tension sur le résistor. Un courant y circule pendant un certain temps. L'énergie électrique consommée par le résistor est de 350 kJ.

Pendant ce temps, la température des 1000 grammes d'eau passe de 15°C à 90°C.

Calculez la quantité d'énergie électrique qui ne se trouve pas sous forme calorifique dans l'eau.

Indiquez toutes les étapes de votre démarche.

25. Vous disposez d'un litre d'une solution d'iodure de potassium, KI, dont la concentration est de 25 g/l.

Avec cette solution, on veut préparer un litre d'une solution de KI dont la concentration sera de 15 g/l.

Écrivez le protocole de manipulation permettant d'obtenir cette nouvelle solution (15 g/l) en justifiant à chaque étape les quantités de solution de KI et d'eau distillée utilisées.

Indiquez toutes les étapes de votre démarche.



ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 3

Les définitions du domaine en sciences

GUIDE DU PARTICIPANT

ACTIVITÉ 3

Les définitions du domaine en sciences

1. Les thèmes
2. Les habiletés et leur limite
3. Les dimensions
4. Les items et leur dimension

1. LES THÈMES

Les contenus des programmes d'études sont synthétisés et regroupés à l'aide de différents thèmes.

2. LES HABILÉTÉS ET LEUR LIMITE

Une habileté, en ce qui a trait à la tâche à accomplir, est le scénario le plus probable pour la majorité des élèves. On pourrait parler d'une habileté comme d'une dominante.

En sciences physiques, on a défini quatre habiletés. Trois de ses quatre habiletés ne sont pas exclusives en soi.

Cette limite aura des répercussions sur les dimensions.

CARACTÉRISER :

Préciser des faits, des phénomènes de l'environnement, une procédure et des techniques, dans le contexte des cheminements d'apprentissage suggérés par le programme.

RELIER :

Établir des rapports entre des faits et des phénomènes, dans le contexte des cheminements d'apprentissage suggérés par le programme.

RÉSOUTRE DES PROBLÈMES EN CLASSE :

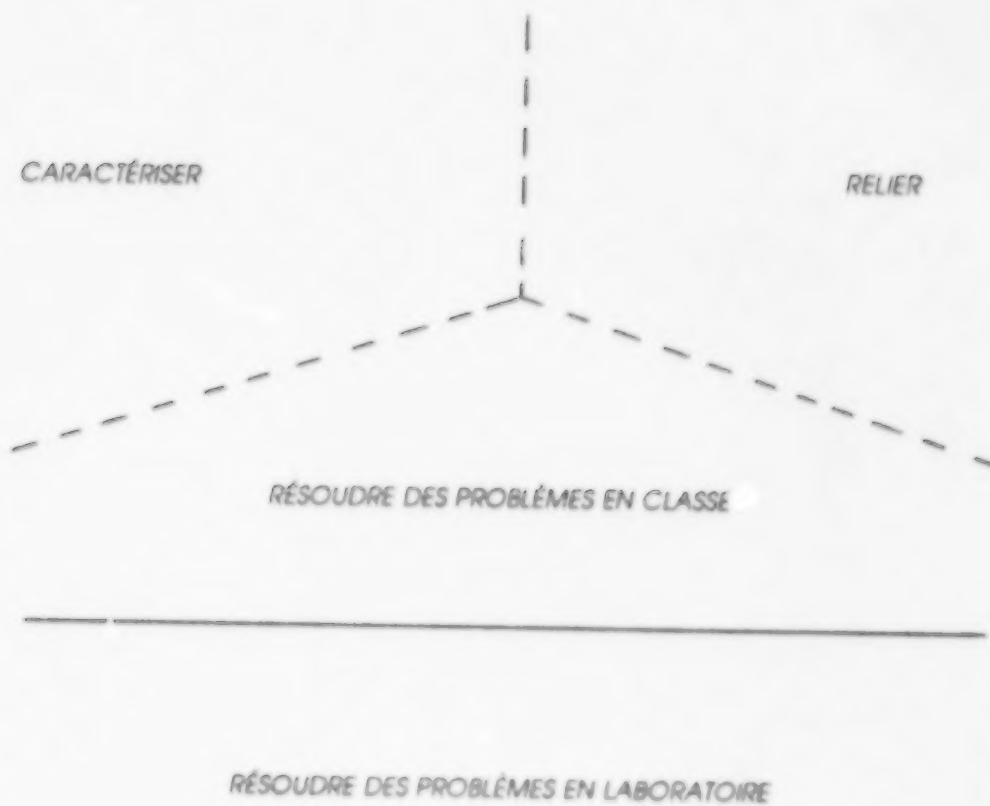
Organiser les composantes de la démarche de résolution de problèmes ou de la méthode scientifique et de l'apprentissage pour solutionner des problèmes en classe.

RÉSOUTRE DES PROBLÈMES EN LABORATOIRE :

Organiser les composantes de la méthode (démarche) scientifique et de l'apprentissage pour solutionner des problèmes en laboratoire.

Note : *S'en tenir aux définitions des habiletés à l'intérieur des définitions du domaine.*

LES HABILETÉS ET LEUR LIMITE



3. LES DIMENSIONS

Les dimensions sont obtenues par le croisement d'une habileté et d'un thème.

Ces dimensions ne sont pas exclusives en soi.

4. LES ITEMS ET LEUR DIMENSION

Exercice sur les items et leur dimension _____



**ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE**

ACTIVITÉ 4

**La construction d'épreuves répondant aux exigences des qualités de validité
et de fidélité**

GUIDE DE L'ANIMATEUR

ACTIVITÉ 4

La construction d'épreuves répondant aux exigences des qualités de validité et de fidélité.

1. *La définition du domaine*
2. *La spécification de l'épreuve*
3. *L'approche du MEQ*
4. *L'approche de l'enseignant*
5. *La rédaction des items*
 - 5.1 *Les composantes de l'item*
 - 5.2 *Les types d'items*
 - 5.3 *Exemples d'items mesurant les habiletés*
 - 5.4 *Les éléments essentiels d'un bon item*
6. *Le montage de l'épreuve*

1. LA DÉFINITION DU DOMAINE

On a vu l'importance de la VALIDITÉ d'une épreuve lorsqu'on se préoccupe de bâtir des instruments de qualité.

Rappelons que la définition du domaine constitue un moyen efficace d'obtenir une synthèse des éléments essentiels d'un programme d'études pour des fins d'évaluation sommative.

La définition du domaine vise essentiellement à assurer la congruence entre un programme d'études et tout instrument de mesure devant servir à une évaluation sommative.

Elle nous fournit quelques règles ou principes à respecter au moment de la préparation d'une l'épreuve, ce qui permet de cerner le contexte dans lequel nous observerons les habiletés de l'élève.

La définition du domaine est donc le PREMIER outil qui nous aide à concevoir une épreuve valide.

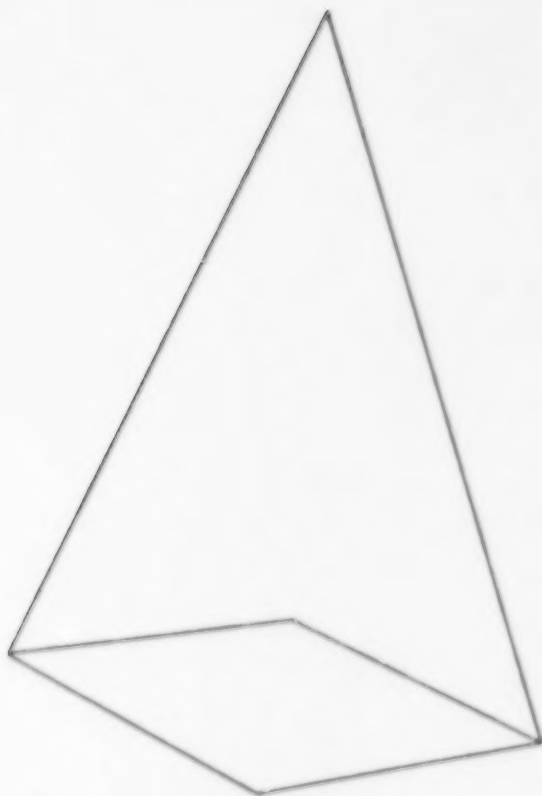
Elle s'adresse au personnel enseignant ainsi qu'aux personnes qui rédigent l'épreuve.

ACÉTATE 4-1 (à juxtaposer sur l'acétate B-1)

ACÉTATE B-1 (Document de base p. 10)

ÉVALUER

4-1



PROGRAMME D'ÉTUDES
(APPRENTISSAGES)

2. LA SPÉCIFICATION DE L'ÉPREUVE

La spécification de l'épreuve est le DEUXIÈME outil pour atteindre la qualité de validité. Elle précise ce qui doit être évalué et comment le faire.

Pour les concepteurs d'épreuves à l'échelle provinciale, cet outil devient essentiel puisqu'il y a plus d'un rédacteur en cause dans l'élaboration de l'épreuve. La spécification de l'épreuve, en précisant davantage le domaine de l'épreuve, aide l'ensemble des rédacteurs à interpréter de la même façon ce qui doit être évalué et les moyens pour le faire.

Elle s'adresse principalement aux personnes qui rédigent l'épreuve, ainsi qu'à celles qui la révisent.

En ayant comme point de repère la définition du domaine, la spécification de l'épreuve précise le but, la nature de l'épreuve et la clientèle cible; elle énumère les objectifs visés et le nombre d'items à l'intérieur des dimensions. De plus, elle précise la durée de l'épreuve, le type d'items à privilégier, les modalités de correction, le matériel autorisé au moment de l'épreuve, les formules et les tables qui seront présentées aux élèves, les graphiques et les symboles utilisés et tout autre renseignement jugé pertinent.

À l'exception des objectifs visés à l'intérieur des dimensions, ces renseignements sont transmis aux enseignants, par l'intermédiaire des commissions scolaires.

Le document d'information concernant la spécification de l'épreuve s'adresse donc au personnel enseignant.

3. L'APPROCHE DU MEQ

Plus le nombre d'élèves à qui s'adresse l'épreuve est grand, plus les concepteurs et les rédacteurs doivent faire des choix contraignants afin de s'assurer que l'épreuve sera valide et fidèle.

L'enseignant qui rédige une épreuve pour son groupe d'élèves a beaucoup moins de contraintes pour s'assurer que l'épreuve soit valide et fidèle. Il lui est possible d'utiliser un plus grand nombre de scénarios se rapprochant même parfois de certains objectifs terminaux; il est le seul correcteur; il peut aisément vérifier si le vocabulaire qu'il utilise est adéquat pour ses élèves, que la longueur de l'épreuve est respectable, que les conditions de passation sont uniformes pour tous et que ses élèves ont déjà rencontré le genre d'items qu'il construit.

À l'opposé, si l'épreuve est administrée à l'échelle nationale, les concepteurs rencontrent plus de contraintes que pour une épreuve à l'échelle provinciale.

Les concepteurs d'épreuves à l'échelle provinciale doivent donc mettre en place des moyens supplémentaires afin de s'assurer que l'épreuve sera aussi valide et fidèle qu'une épreuve interne.

Quelles sont donc les contraintes que rencontrent les concepteurs d'une épreuve à l'échelle provinciale?

Demander aux participants d'associer la contrainte nommée à la qualité qu'elle risque d'affecter, la validité ou la fidélité.

Les contraintes que rencontrent les concepteurs d'une épreuve administrée à l'échelle provinciale.

Contraintes qui risquent d'affecter surtout la validité de l'épreuve :

- Est-ce que les enseignants interprètent le programme d'étude de la MÊME façon?
- Est-ce que les enseignants interprètent les objectifs terminaux de la MÊME façon?
- Est-ce que les enseignants vont interpréter la définition du domaine de la MÊME façon?
- Est-ce que les enseignants interprètent les habiletés de la MÊME façon?
- Est-ce que les items rédigés par un groupe d'enseignants seront clairs et précis pour TOUS les élèves de la province?
- En raison du nombre d'élèves, les concepteurs d'épreuves à l'échelle provinciale DOIVENT s'en tenir au scénario suivant : épreuve papier-crayon, même s'ils ne peuvent évaluer certains objectifs;
- Mais est-ce que l'épreuve papier-crayon est un BON scénario pour évaluer les objectifs du programme d'études?
- En raison de la contrainte «durée de l'épreuve», les concepteurs d'épreuves à l'échelle provinciale NE PEUVENT mesurer tous les objectifs terminaux d'un programme d'études. Ils doivent faire un choix d'objectifs à l'intérieur du programme;
- Les objectifs qui ne seront pas mesurés à l'échelle provinciale, en raison des contraintes, seront-ils MIS DE CÔTÉ par les enseignants?

- Est-ce que les objectifs choisis sont **REPRÉSENTATIFS** de l'ensemble du programme?
- Est-ce que les rédacteurs d'épreuves interprètent le programme d'études de la **MÊME** façon?
- Est-ce que les rédacteurs interprètent les objectifs terminaux de la **MÊME** façon?
- Est-ce que les rédacteurs vont interpréter la définition du domaine de la **MÊME** façon?
- Est-ce que les rédacteurs interprètent les habiletés de la **MÊME** façon?
- Est-ce que la longueur de l'épreuve est **ADÉQUATE** pour mesurer un ensemble d'objectifs représentatifs du programme d'études?
- Est-ce que **TOUS** les élèves seront évalués justement en fonction du programme d'études?
- Autres contraintes.

Contraintes qui risquent d'affecter surtout la fidélité de l'épreuve :

- Le grand nombre de correcteurs : est-ce que **TOUS** les correcteurs de la province corrigent de la même façon et que chaque correcteur corrige **TOUTES** ses copies de la même façon?
- Les concepteurs d'épreuves à l'échelle provinciale n'ont **AUCUN** moyen d'évaluer justement l'élève qui répond au hasard, celui qui triche ou celui qui n'a pu donner son vrai rendement en raison de son état de santé par exemple.
- Est-ce que le vocabulaire et la syntaxe utilisés seront clairs pour **TOUS** les élèves qui devront répondre à l'épreuve?

- Est-ce que les items sont exempts de stéréotypes pour TOUS les élèves?
- Est-ce que TOUS les élèves de la province ont déjà travaillé avec le genre d'items utilisés dans l'épreuve?
- Est-ce que les consignes sont claires, précises et complètes pour TOUS les élèves de la province?
- Est-ce que la longueur de l'épreuve est adéquate pour l'ENSEMBLE des élèves de la province?
- Est-ce que les conditions de passation sont uniformes pour TOUS les élèves de la province?
- Est-ce que la journée durant laquelle on administre l'épreuve est appropriée pour TOUS les élèves de la province?
- Est-ce que le genre d'épreuve utilisé évaluera justement TOUS les élèves?
- Autres contraintes.

Lorsqu'un enseignant rédige une épreuve qui sera administrée à ses élèves, il peut facilement et rapidement répondre adéquatement à la majorité des questions que nous venons de soulever. Il est le seul rédacteur, le seul correcteur et il connaît bien TOUS les élèves à qui s'adresse l'épreuve.

Pour que le MEQ puisse concevoir des épreuves à l'échelle provinciale valides et fidèles, il doit donc recourir à des outils supplémentaires comme la définition du domaine et la spécification de l'épreuve, et ce, avant même qu'il pense à analyser son épreuve en fonction des pistes fournies pour que les items et l'ensemble de l'épreuve soient considérés comme valides et fidèles.

De plus, il doit arrêter son choix sur une grille de correction simple et précise afin que l'ensemble des correcteurs de la province l'interprète de la même façon.

Le cadre général de l'élaboration des épreuves uniques décrit l'ensemble des opérations et des outils mis en place par le MEQ afin d'obtenir des épreuves de qualité.

Les principales étapes sont les suivantes :

1. la définition du domaine;
2. la spécification de l'épreuve;
3. la production des items;
4. la vérification de l'épreuve et la traduction;
 - sur le plan du contenu et de la docimologie,
 - sur le plan linguistique,
 - la mise à l'essai;
5. le montage de l'épreuve;
6. la composition typographique.

Exercice : associer les qualités des instruments de mesure aux principales étapes de la préparation des épreuves à l'échelle provinciale.

Le but de l'exercice est de faire ressortir les raisons de chacune des étapes, autrement dit ce en quoi chacune des étapes améliore la qualité des épreuves.

Il est certain que les deux qualités sont liées; demander aux participants de déterminer la principale qualité associée à chaque étape.

ACÉTATES 4-2 et 4-2s

Référence :

MEQ, *Cadre général d'élaboration des épreuves uniques*, 1986, 14 p.

VALIDITÉ ET FIDÉLITÉ

Associez les qualités VALIDITÉ et FIDÉLITÉ aux principales étapes de la préparation des épreuves à l'échelle provinciale. Inscrivez le nom de la qualité vis-à-vis les étapes. Il se peut qu'une étape vise à la fois à améliorer la validité et la fidélité de l'épreuve.

Les principales étapes de la rédaction des épreuves à l'échelle provinciale sont les suivantes :

1. la définition du domaine; _____
2. la spécification de l'épreuve; _____
3. la production des items; _____
4. la vérification de l'épreuve et la traduction; _____
 - a) sur le plan du contenu et de la docimologie, _____
 - b) sur le plan linguistique, _____
 - c) la mise à l'essai; _____
5. le montage de l'épreuve; _____
6. la composition typographique. _____

VALIDITÉ ET FIDÉLITÉ

Associez les qualités VALIDITÉ et FIDÉLITÉ aux principales étapes de la préparation des épreuves à l'échelle provinciale. Inscrivez le nom de la qualité vis-à-vis les étapes. Il se peut qu'une étape vise à la fois à améliorer la validité et la fidélité de l'épreuve.

Les principales étapes de la rédaction des épreuves à l'échelle provinciale sont les suivantes :

1. la définition du domaine; __VALIDITÉ__

2. la spécification de l'épreuve; VALIDITÉ-FIDÉLITÉ

3. la production des items;

4. la vérification de l'épreuve et la traduction;
 - a) sur le plan du contenu __VALIDITÉ__
 et de la docimologie, __FIDÉLITÉ__
 - b) sur le plan linguistique, __FIDÉLITÉ__
 - c) la mise à l'essai; __FIDÉLITÉ__

5. le montage de l'épreuve; __FIDÉLITÉ__

6. la composition typographique.

4. L'APPROCHE DE L'ENSEIGNANT

Pour l'enseignant, la rédaction d'une épreuve sommative valide et fidèle comporte moins d'étapes. On peut les résumer comme suit :

1. en partant de la définition du domaine ou d'un tableau de spécifications, l'enseignant précise les objectifs et les habiletés qui seront évalués;

(Cette étape améliore la validité de votre épreuve.)

2. il produit les items en fonction des objectifs et des habiletés qu'il a prédéterminés;

3. il vérifie l'épreuve :

- en s'assurant de la congruence des items avec les objectifs et les habiletés,

(Cette étape améliore la validité de votre épreuve.)

- en s'assurant que les items soient bien construits,

(Cette étape améliore la validité et la fidélité de votre épreuve.)

- en examinant l'aspect linguistique;

(Cette étape améliore la fidélité de votre épreuve.)

4. enfin, il fait le montage de l'épreuve.

5. LA RÉDACTION D'ITEMS

5.1 Les composantes de l'item

La mise en situation *Elle fournit les circonstances générales de l'item à partir de contextes.*

Les données *Les données du problème.*

La tâche à accomplir *Elle présente ce que l'élève doit faire pour que l'on puisse juger de son habileté ou de ses connaissances.*

La consigne *Elle indique à l'élève comment répondre.*

5.2 Les types d'items

Les principaux types d'items utilisés en sciences sont les suivants :

- Item à réponse choisie,
- Item de type appariement,
- Item à réponse courte,
- Item à réponse développée.

Vous trouverez les principes de rédaction propres à chaque type d'item à l'intérieur du GUIDE 2 (G-2).

Il faut se rappeler que la nature de l'objectif et l'habileté à mesurer sont plus importantes que le type d'item à utiliser. Choisissez le type d'item en fonction de l'objectif et de l'habileté.

Parfois, nous éprouvons de la difficulté à rédiger un item. Bien souvent, en changeant simplement le type d'item, nous en facilitons la rédaction.

5.3 Exemples d'items mesurant les habiletés

SCIENCES PHYSIQUES **HABILETÉ-Résoudre des problèmes en classe**

OBJECTIF :

Déterminer la quantité d'énergie électrique consommée dans un circuit au moyen...

ITEM

La chaleur dans une couveuse électrique est produite par une ampoule de 40 watts. L'ampoule est branchée au secteur dont la différence de potentiel est de 120 volts. Les oeufs ont besoin de 21 jours pour éclore. Hydro-Québec exige 0,08 \$/kWh.

Quel est le coût de l'énergie électrique utilisée?

SOLUTION

Item à réponse choisie : les leurres et la bonne réponse.

Item à réponse construite :

réponse courte : réponse attendue et autres réponses acceptées.

réponse développée : démarche et résultat.

OBJECTIF :

Identifier, dans des situations problèmes proposées, les données utiles à leur résolution.

ITEM

La chaleur dans une couveuse électrique est produite par une ampoule de 40 watts. L'ampoule est branchée au secteur dont la différence de potentiel est de 120 volts. Les oeufs ont besoin de 21 jours pour éclore. Hydro-Québec exige 0,08 \$/kWh.

Quelles sont les données utiles pour calculer l'énergie électrique consommée?

SOLUTION

Item à réponse choisie : les leurres et la bonne réponse.

Item à réponse construite :

réponse courte : réponse attendue et autres réponses acceptées.

réponse développée : démarche et résultat.

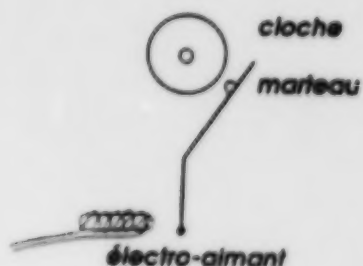
Toutes les données sont utiles excepté le prix par kWh.

OBJECTIF :

Analyser des caractéristiques d'un champ magnétique mises en évidence lors de ses travaux scientifiques.

ITEM

Pourquoi dans la sonnerie illustrée ci-dessous utilise-t-on un électro-aimant?



SONNERIE

SOLUTION

Item à réponse choisie : les leurres et la bonne réponse.

Item à réponse construite :

réponse courte : réponse attendue et autres réponses acceptées.

réponse développée : démarche et résultat.

Parce que nous pouvons contrôler le champ magnétique.

EXERCICE 1

À partir de la banque d'items présentés à l'activité 3, demander aux participants d'associer un item à son habileté.

EXERCICE 2

Rédaction d'items à partir de la définition du domaine.

Par groupe de trois ou quatre, les participants rédigent quelques items :

1. *Chaque groupe choisit un objectif à l'intérieur du contenu.*

On suggère que chaque groupe choisisse un objectif différent. À la fin de l'activité, on pourrait produire une épreuve relativement complète.

2. *Chaque groupe rédige des items.*
3. *Chaque groupe transcrit sur des acétates l'objectif qu'il a retenu, l'habileté, et les items rédigés.*

ACÉTATES 4-3 A) et 4-3 B)

4. *À tour de rôle, chaque groupe présente ses items.*

5. L'animateur invite tous les participants à améliorer, s'il y a lieu, les items présentés en se prononçant sur les critères de congruence au regard de l'objectif et de l'habileté, et sur les règles de rédaction du type d'item choisi.

Les participants peuvent s'aider des GUIDES 1 et 2 ou du guide intitulé SYNTHÈSE DE L'ITEM présenté au point 5.4.

Assemblée plénière

L'objectif prime.

L'évaluation est au service de la matière à l'étude et non l'inverse.

Évaluation «assomative»!

SCIENCES PHYSIQUES HABILITÉ _____

OBJECTIF : _____

ITEM

Mise en situation _____
Données _____
Tâche _____
Consigne _____

SOLUTION

Item à réponse choisie : les leures et la bonne réponse.

Item à réponse construite :

réponse courte : réponse attendue et autres réponses acceptées.

réponse développée : démarche et résultat.

5.4 Les éléments essentiels d'un bon item

Pour appuyer sur les éléments essentiels d'un bon item, nous vous présentons le guide SYNTHÈSE DE L'ITEM.

À l'intérieur de ce guide, fait à partir du GUIDE 1, nous avons reproduit les éléments les plus importants nous permettant de nous prononcer sur la qualité d'un item. De plus, vous pouvez consulter le GUIDE 2 qui contient les principes de rédaction de chaque type d'item.

SYNTHÈSE DE L'ITEM

POUR UN ITEM DE QUALITÉ, ASSUREZ-VOUS :

- | | | |
|----|--|--------------------------|
| 1- | que le choix de l'objectif est pertinent, | ☐ |
| 2- | de la congruence de l'item avec l'objectif
et l'habileté mesurés, | ☐ |
| 3- | que les consignes sont claires, précises
et complètes, | ☐ |
| 4- | que les élèves ont déjà une connaissance du
type d'item utilisé, | ☐ |
| 5- | que le type d'item utilisé est approprié
pour mesurer l'objectif au regard d'une
habileté, | ☐ |
| 6- | que l'item est exempt de tout
stéréotype, | ☐ |
| 7- | que le nombre de points accordés à chaque
item est indiqué, | ☐ |
| 8- | que les principes de rédaction propres
à chaque item sont respectés. | ☐ |

6. LE MONTAGE DE L'ÉPREUVE

Le MEQ groupe ses items par type d'item et par thème ou par habileté. Il tient également compte de la difficulté.

Vous pouvez grouper vos items de la même façon ou choisir tout autre montage qui facilite le travail de l'élève sans causer d'ambiguïté.

Pour appuyer sur les éléments essentiels d'une épreuve valide et fidèle, nous vous présentons le guide intitulé SYNTHÈSE DE L'ÉPREUVE.

À l'intérieur de ce guide, fait à partir du GUIDE 3, nous avons reproduit les éléments les plus importants des qualités de validité et de fidélité vous permettant de vous prononcer rapidement sur la qualité d'une épreuve.

SYNTHÈSE DE L'ÉPREUVE

POUR UNE ÉPREUVE VALIDE, ASSUREZ-VOUS :

- 1- que l'ensemble des objectifs et des habiletés que vous mesurez sont représentatifs d'une partie du programme d'études ou du programme d'études en entier, ☐
- 2- qu'il y a congruence entre vos items et vos objectifs et habiletés. ☐

POUR UNE ÉPREUVE FIDÈLE, ASSUREZ-VOUS MAINTENANT :

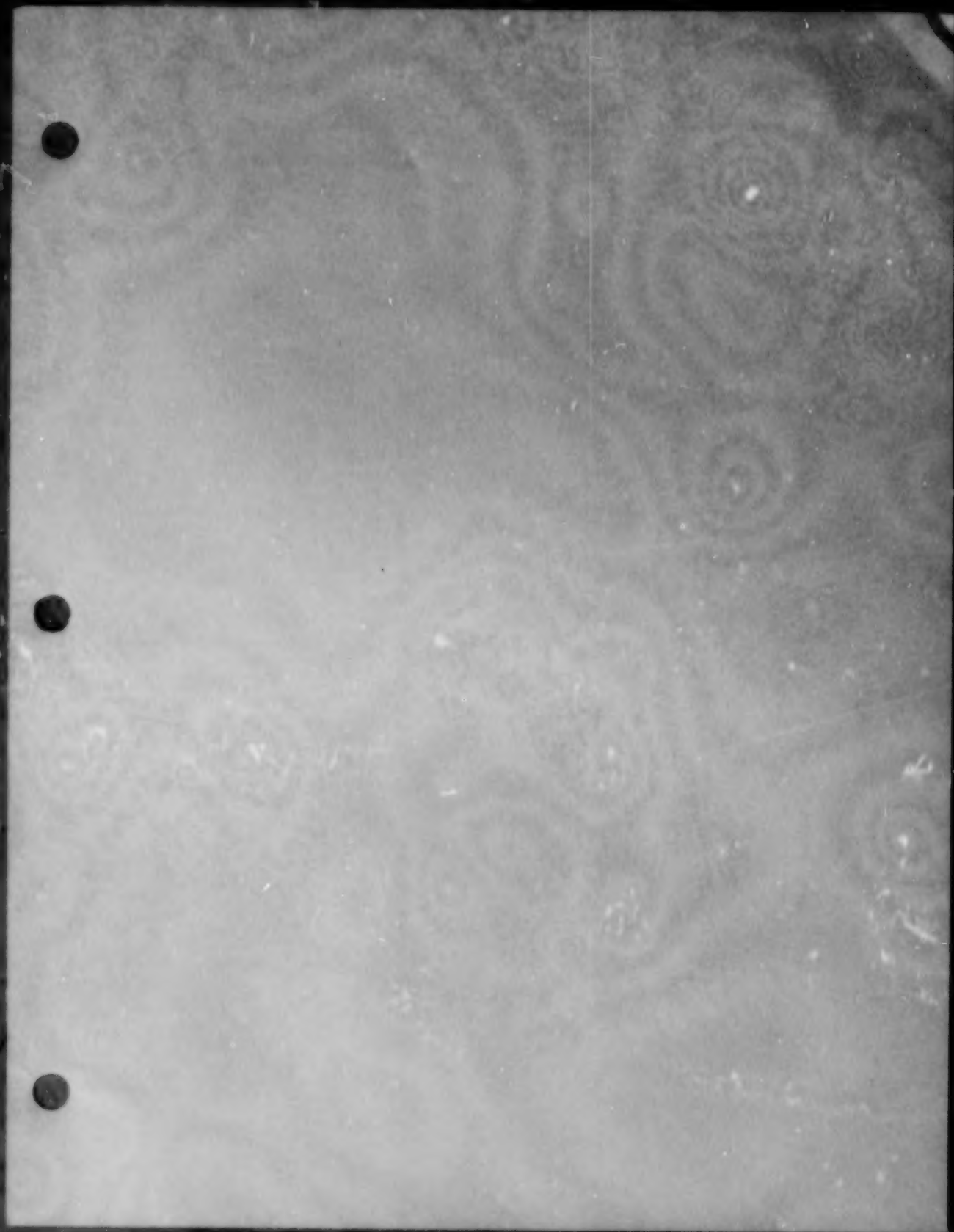
- 1- que la présentation est claire et lisible, ☐
- 2- que l'épreuve est exempte d'erreurs grammaticales, orthographiques et typographiques, ☐
- 3- que les directives de l'épreuve sont complètes et bien en évidence, ☐
- 4- que chaque item apparaît sur une même page, ☐
- 5- que chaque item est indépendant des autres items, ☐
- 6- que les critères de correction sont clairs, ☐
- 7- que la pondération des items représente bien l'importance des objectifs et des habiletés à l'intérieur du programme d'études, ☐
- 8- que la longueur de l'épreuve est appropriée par rapport au temps alloué, ☐
- 9- que les conditions de passation seront les mêmes pour tous. ☐

Vous trouverez, dans les pages qui suivent, les guides suivants :

GUIDE 1 : GRILLE D'ANALYSE DE L'ITEM

GUIDE 2 : LES PRINCIPES DE RÉDACTION DE CHAQUE TYPE
D'ITEM

GUIDE 3 : GUIDE DE RÉVISION DE L'ENSEMBLE DE L'ÉPREUVE



GRILLE D'ANALYSE DE L'ITEM

1. LE CHOIX DE L'OBJECTIF EST-IL PERTINENT?
(LA PERTINENCE) ☐
2. EST-CE QUE L'ITEM CORRESPOND À L'OBJECTIF
DONT IL DOIT MESURER L'ATTEINTE? (LA CONGRUENCE) ☐
3. LES CONNAISSANCES OU LES HABILITÉS QUE L'ITEM MESURE
SONT-ELLES SUFFISAMMENT IMPORTANTES POUR QU'ON LES MESURE? ☐
4. LA MISE EN SITUATION EST-ELLE CONCISE ET PERTINENTE? ☐
5. LA TÂCHE EST-ELLE PRÉCISE ET CLAIRE? ☐
6. LES DONNÉES SONT-ELLES ADAPTÉES AU CONTEXTE? ☐
7. LES CONSIGNES SONT-ELLES PRÉCISES, CLAIRES ET COMPLÈTES?
(LE NOMBRE DE BONNES RÉPONSES S'IL Y EN A PLUS D'UNE,
L'ENDROIT OU L'ÉLÈVE DOIT INDiquer SA RÉPONSE ET LA
FAÇON DE L'EXPRIMER, L'ENDROIT OU L'ÉLÈVE EXÉCUTE SES
CALCULS, ETC.) ☐
8. LE TYPE D'ITEM EST-IL APPROPRIÉ POUR MESURER L'OBJECTIF
RETENU? ☐
9. LE VOCABULAIRE EST-IL ADAPTÉ AUX ÉLÈVES? ☐
10. LES RÈGLES DE LA SYNTAXE SONT-ELLES RESPECTÉES? ☐
11. LES ÉLÈVES LE CONNAISSENT-ILS LE TYPE D'ITEM UTILISÉ? ☐
12. LES RENSEIGNEMENTS SUPERFLUS DONT LA PRÉSENCE RISQUE
DE COMPLIQUER INUTILEMENT LA TÂCHE DE L'ÉLÈVE SONT-ILS
ÉLIMINÉS? ☐

GRILLE D'ANALYSE DE L'ITEM
SUITE

- 13. L'ITEM EST-IL EXEMPT DE TOUTE ABERRATION SUSCEPTIBLE DE COMPROMETTRE LES RÉSULTATS? ☐
- 14. L'ITEM EST-IL EXEMPT DE TOUT STÉRÉOTYPE CULTUREL, ETHNIQUE, SOCIAL ET SEXISTE? ☐
- 15. L'ITEM EST-IL EXEMPT DE PHRASES TIRÉES DE MANUELS SCOLAIRES OU DE NOTES DE COURS? ☐
- 16. LA SOURCE DES CITATIONS EST-ELLE INDICUÉE? ☐
- 17. LE NOMBRE DE POINTS ACCORDÉS À CHAQUE ITEM EST-IL INDICUÉE? ☐
- 18. CHAQUE ITEM APPARAÎT-IL SUR UNE MÊME PAGE? ☐
- 19. LES PRINCIPES DE RÉDACTION PROPRES À CHAQUE TYPE D'ITEM SONT-ILS RESPECTÉS? ☐

**LES PRINCIPES DE RÉDACTION PROPRES À CHAQUE TYPE
D'ITEM SONT-ILS RESPECTÉS?**

ITEM À RÉPONSE CHOISIE

1. L'ÉNONCÉ CONTIENT TOUTES LES LIMITES ET LES CONDITIONS
NÉCESSAIRES À LA RÉOLUTION DU PROBLÈME. ☐
2. UTILISER LA PHRASE IMPÉRATIVE OU INTERROGATIVE. ☐
3. SÉPARER LA CONSIGNE DE L'ÉNONCÉ. ☐
4. PLACER DANS L'ÉNONCÉ TOUS LES MOTS QUI AUTREMENT
SE RÉPÉTERAIENT DANS CHAQUE RÉPONSE SUGGÉRÉE. ☐
5. CHAQUE RÉPONSE SUGGÉRÉE DOIT ÊTRE CONSISTANTE ET
APPROPRIÉE À L'ÉNONCÉ (À PEU PRÈS DE MÊME LONGUEUR
DE MÊME COMPLEXITÉ ET AYANT LE MÊME ATTRAIT). ☐
6. LES RÉPONSES SUGGÉRÉES DOIVENT ÊTRE CONCISES ET
SIMPLES (À MOINS QUE LA NATURE DU PROBLÈME À
RÉSOUTRE NE PRESCRIVE LE CONTRAIRE). ☐
7. AUTANT QUE POSSIBLE, LES RÉPONSES SUGGÉRÉES DOIVENT
ÊTRE PLACÉES EN COLONNE PLUTÔT QU'EN TANDEM. ☐
8. IL EST PLUS FACILE DE LIRE LES RÉPONSES QUI SONT
PLACÉES EN COLONNE PLUTÔT QU'EN TANDEM. ☐
9. ÉVITER LES NÉGATIONS AMBIGUËS. ☐
10. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
11. INDiquer LE NOMBRE DE POINTS ACCORDÉS. ☐

ITEM DE TYPE VRAI OU FAUX

1. PRÉSENTER UNE SEULE IDÉE IMPORTANTE. ☐
2. PRÉSENTER DES ITEMS MANIFESTEMENT VRAIS OU FAUX. ☐
3. PRÉSENTER DES ITEMS PLUTÔT BREFS QUE LONGS. ☐
4. ÉTOFFER SUFFISAMMENT L'ÉNONCÉ POUR QUE L'ÉLÈVE
DISPOSE DES CRITÈRES NÉCESSAIRES À LA RÉOLUTION
DU PROBLÈME ☐
5. ÉVITER LES EXTRAITS TEXTUELS. ☐
6. UTILISER AVEC PRUDENCE LES TERMES - TOUS - TOUJOURS
JAMAIS, QUI SONT ASSOCIÉS À DES ÉNONCÉS FAUX; LES
TERMES - PARFOIS - SOUVENT - LA PLUPART, QUI SONT
ASSOCIÉS À DES ÉNONCÉS VRAIS ET LES TERMES - FRÉQUENT
NOMBREUX - UNE FOULE - PEU, QUI SONT AMBIGUS. ☐
7. IL EST SOUHAITABLE DE PLACER DANS UNE ÉPREUVE PLUS
D'ITEMS FAUX. ☐
8. ÉVITER LES NÉGATIONS AMBIGUËS. ☐
9. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
10. INDIQUER LE NOMBRE DE POINTS ACCORDÉS. ☐

ITEM DE TYPE APPARIEMENT

1. DONNER LE SCHÉMA DE BASE DE L'ASSOCIATION ET DÉCRIRE CHAQUE ENSEMBLE. ☐
2. INDIQUER S'IL Y A OU NON RÉPÉTITION DES ÉLÉMENTS DE L'ENSEMBLE-RÉPONSES. ☐
3. PLACER DANS CHAQUE ENSEMBLE DES ÉLÉMENTS HOMOGENES. ☐
4. NE PAS COMPTER AU TOTAL PLUS DE DIX ÉLÉMENTS DANS L'ENSEMBLE-RÉPONSES. (SIX POUR LE PRIMAIRE) ☐
5. PLACER UN PLUS GRAND NOMBRE D'ÉLÉMENTS DANS L'ENSEMBLE-RÉPONSES. ☐
6. DÉSIGNER L'ENSEMBLE-STIMULI PAR DES NOMBRES ET L'ENSEMBLE-RÉPONSES PAR DES LETTRES. ☐
7. HABITUELLEMENT L'ENSEMBLE-STIMULI EST PLACÉ À GAUCHE. ☐
8. PLACER LES ÉLÉMENTS DE L'ENSEMBLE-RÉPONSES DANS UN ORDRE QUI FACILITE LE TRAVAIL DE L'ÉLÈVE. ☐
9. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
10. INDIQUER LE NOMBRE DE POINTS ACCORDÉS. ☐

ITEM DE TYPE PHRASE À COMPLÉTER

1. ON NE DOIT ENLEVER QUE LES TERMES IMPORTANTS DE LA PHRASE. ☐
2. NE PLACER QU'UNE OU DEUX OMISSIONS PAR PHRASE À COMPLÉTER. ☐
3. PLACER LES OMISSIONS VERS LA FIN DE LA PHRASE. ☐
4. LA LONGUEUR DE L'ESPACE DOIT CORRESPONDRE À LA LONGUEUR DU MOT OMIS. ☐
5. SI LA RÉPONSE PORTE SUR DES QUANTITÉS, INDiquer LES UNITÉS DEMANDÉES ET LE DEGRÉ DE PRÉCISION. (SI DÉSIRÉ) ☐
6. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
7. INDiquer LE NOMBRE DE POINTS ACCORDÉS. ☐

**ITEM À RÉPONSE CONSTRUITE
(RÉPONSE COURTE)**

1. IL EST PRÉFÉRABLE DE PRÉSENTER LE PROBLÈME EN
POSANT UNE QUESTION DIRECTE PLUTÔT QU'EN
FORMULANT UNE AFFIRMATION. ☐
2. ÉVITER LES EXPRESSIONS TEXTUELLES. ☐
3. IL FAUT PRÉVOIR UN ESPACE APRÈS CHAQUE ITEM POUR
QUE L'ÉLÈVE PUISSE ÉCRIRE SA RÉPONSE IMMÉDIATEMENT
À MOINS QU'IL Y AIT UNE FEUILLE-RÉPONSES. ☐
4. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
5. INDiquer LE NOMBRE DE POINTS ACCORDÉS. ☐

**ITEM À RÉPONSE CONSTRUITE
(RÉPONSE DÉVELOPPÉE)**

1. IL FAUT STRUCTURER, ENCADRER SUFFISAMMENT LE PROBLÈME PRÉSENTÉ DANS L'ITEM POUR QUE LA RÉPONSE SOIT ORIENTÉE VERS CE QUE L'ON SE PROPOSE DE MESURER. ☐
2. MIEUX VAUT INDiquer L'ORDRE DE GRANDEUR DE LA RÉPONSE SOUHAITÉE. ☐
3. PRÉVOIR L'ESPACE NÉCESSAIRE POUR LA SOLUTION. ☐
4. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
5. INDiquer LE NOMBRE DE POINTS ACCORDÉS ET, SI NÉCESSAIRE, LE TEMPS NORMALEMENT ALLOUÉ POUR RÉSOUDRE LE PROBLÈME. ☐
6. SI L'ENSEIGNANT LE JUGE À PROPOS, IL PEUT AVERTIR L'ÉLÈVE QU'IL SERA PÉNALISÉ S'IL INTRODUIT DANS SA RÉPONSE DES ÉLÉMENTS ÉTRANGERS AU PROBLÈME À RÉSOUDRE. ☐

GUIDE DE RÉVISION D'UNE ÉPREUVE

63

POUR L'ENSEMBLE DE L'ÉPREUVE

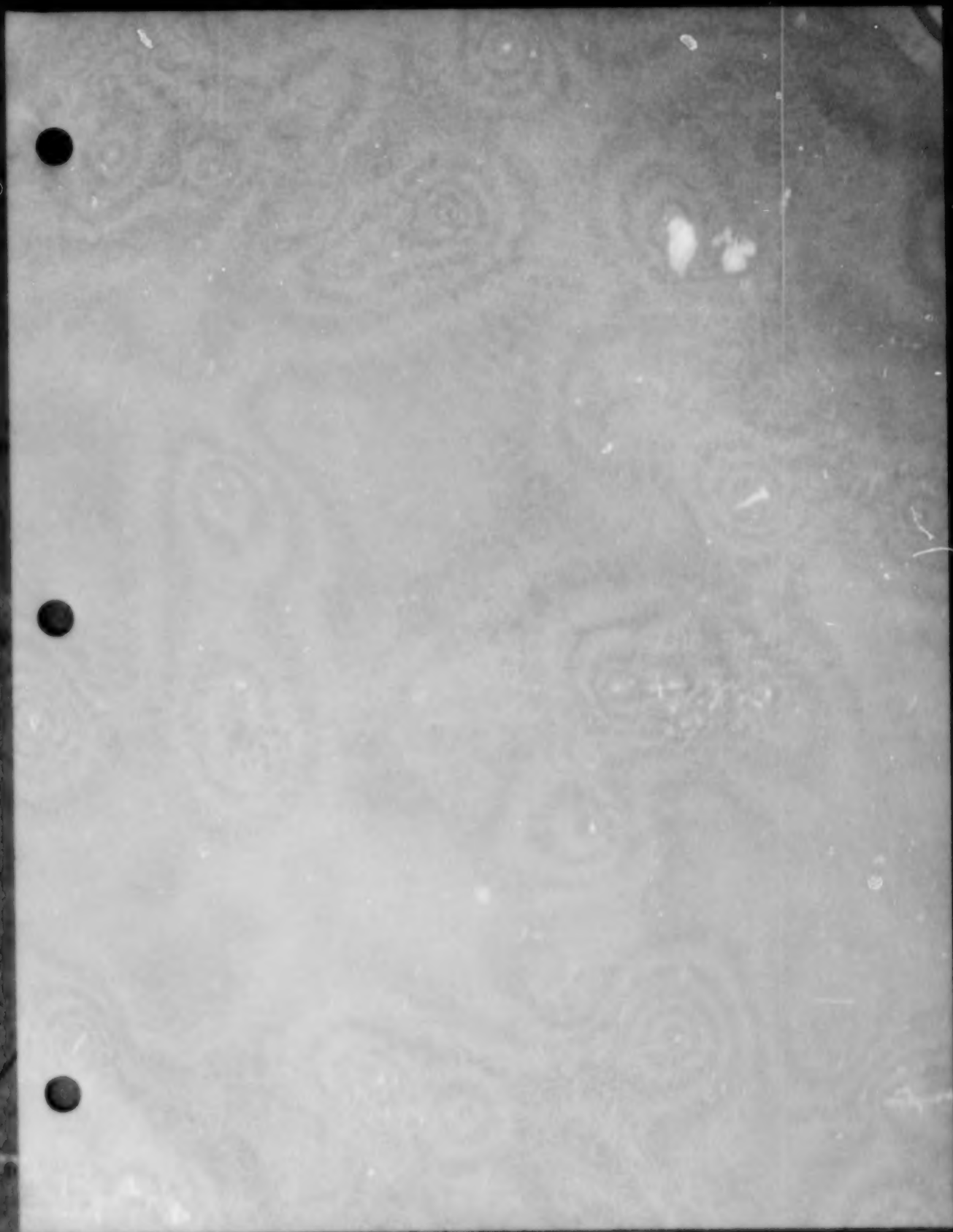
1. LE BUT DE L'ÉPREUVE EST-IL CLAIREMENT ÉTABLI? ☐
2. Y-A-IL CONGRUENCE ENTRE LE CONTENU DE L'ÉPREUVE ET LES OBJECTIFS QUE L'ON VEUT ÉVALUER? ☐
3. LE DEGRÉ DE DIFFICULTÉ DE L'ÉPREUVE EST-IL APPROPRIÉ, ÉTANT DONNÉ LE BUT DE L'ÉVALUATION ET LE NIVEAU DES ÉLÈVES? ☐
4. LE CHOIX DU TYPE D'ITEM EST-IL JUDICIEUX :
LE TYPE D'ITEM EST-IL APPROPRIÉ POUR MESURER L'OBJECTIF RETENU? ☐
5. L'AGENCEMENT DES ITEMS EST-IL PERTINENT?
(PAR TYPE D'ITEM, PAR ORDRE DE DIFFICULTÉ OU PAR ORDRE DE PRÉSENTATION DE LA MATIÈRE) ☐
6. LA LONGUEUR DE L'ÉPREUVE EST-ELLE APPROPRIÉE PAR RAPPORT AU TEMPS ALLOUÉ, AU NOMBRE D'ITEMS ET AU NIVEAU DES ÉLÈVES? ☐
7. LES DIRECTIVES SONT-ELLES CLAIRES, COMPLÈTES, PRÉCISES ET BIEN EN ÉVIDENCE?
(LE TEMPS ALLOUÉ, L'ENDROIT OÙ L'ÉLÈVE DOIT INSCRIRE SA RÉPONSE ET LA FAÇON DE L'EXPRIMER, LE NOMBRE DE BONNE RÉPONSES S'IL Y EN A PLUS D'UNE, L'ENDROIT OÙ L'ÉLÈVE EXÉCUTE SES CALCULS, LA VALEUR DE L'ÉPREUVE, ETC.) ☐
8. LES CONSIGNES POUR RÉPONDRE AUX ITEMS SONT-ELLES CLAIRES ET RÉDIGÉES AVEC UN VOCABULAIRE ACCESSIBLE AUX ÉLÈVES? ☐
9. LES TEXTES UTILISÉS SONT-ILS ADAPTÉS AUX INTÉRÊTS ET AU NIVEAU DES ÉLÈVES? ☐

**POUR L'ENSEMBLE DE L'ÉPREUVE
SUITE**

- 10. LE NIVEAU DE LANGUE (VOCABULAIRE ET SYNTAXE)
EST-IL ADAPTÉ AUX ÉLÈVES? ☐
- 11. CHAQUE ITEM EST-IL INDÉPENDANT DES AUTRES ITEMS? ☐
- 12. L'ÉPREUVE EST-ELLE EXEMPTÉ D'ERREURS GRAMMATICALES,
ORTHOGRAPHIQUES ET TYPOGRAPHIQUES? ☐
- 13. LES ILLUSTRATIONS SONT-ELLES CLAIRES? ☐
- 14. LES CARACTÈRES D'IMPRIMERIE SONT-ILS LISIBLES? ☐

POUR LA CLÉ DE CORRECTION

1. LA RÉPONSE EST-ELLE CLAIREMENT INDIQUÉE POUR
CHAQUE ITEM? ☐
2. Y A-T-IL D'AUTRES RÉPONSES ACCEPTABLES POUR
CHAQUE ITEM? ☐
3. LES AUTRES RÉPONSES ACCEPTABLES SONT-ELLES INDIQUÉES? ☐
4. LE NOMBRE DE POINTS ACCORDÉS À CHAQUE ITEM OU ASPECT
DE L'ITEM EST-IL INDIQUÉ? ☐
5. LES CRITÈRES DE CORRECTION POUR CHAQUE ITEM À RÉPONSE
CONSTRuite SONT-ILS INDIQUÉS? ☐
6. LES CRITÈRES DE CORRECTION SONT-ILS CLAIRS? ☐



ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 4

*La construction d'épreuves répondant aux exigences des qualités de validité
et de fidélité*

GUIDE DU PARTICIPANT

ACTIVITÉ 4

La construction d'épreuves répondant aux exigences des qualités de validité et de fidélité

1. La définition du domaine
2. La spécification de l'épreuve
3. L'approche du MEQ
4. L'approche de l'enseignant
5. La rédaction des items
 - 5.1 Les composantes de l'item
 - 5.2 Les types d'items
 - 5.3 Exemples d'items mesurant les habiletés
 - 5.4 Les éléments essentiels d'un bon item
6. Le montage de l'épreuve

1. LA DÉFINITION DU DOMAINE

Rappelons que la définition du domaine constitue un moyen efficace d'obtenir une synthèse des éléments essentiels d'un programme d'études pour des fins d'évaluation sommative.

La définition du domaine vise essentiellement à assurer la congruence entre un programme d'études et tout instrument de mesure devant servir à une évaluation sommative.

Elle nous fournit quelques règles ou principes à respecter au moment de la préparation d'une l'épreuve, ce qui permet de cerner le contexte dans lequel nous observerons les habiletés de l'élève.

La définition du domaine est donc le PREMIER outil qui nous aide à concevoir une épreuve valide.

Elle s'adresse au personnel enseignant ainsi qu'aux personnes qui rédigent l'épreuve.

2. LA SPÉCIFICATION DE L'ÉPREUVE

La spécification de l'épreuve est le DEUXIÈME outil pour atteindre la qualité de validité. Elle précise ce qui doit être évalué et comment le faire.

Pour les concepteurs d'épreuves à l'échelle provinciale, cet outil devient essentiel puisqu'il y a plus d'un rédacteur en cause dans l'élaboration de l'épreuve. La spécification de l'épreuve, en précisant davantage le domaine de l'épreuve, aide l'ensemble des rédacteurs à interpréter de la même façon ce qui doit être évalué et les moyens pour le faire.

Elle s'adresse principalement aux personnes qui rédigent l'épreuve, ainsi qu'à celles qui la révisent.

En ayant comme point de repère la définition du domaine, la spécification de l'épreuve précise le but, la nature de l'épreuve et la clientèle cible; elle énumère les objectifs visés et le nombre d'items à l'intérieur des dimensions. De plus, elle précise la durée de l'épreuve, le type d'items à privilégier, les modalités de correction, le matériel autorisé au moment de l'épreuve, les formules et les tables qui seront présentées aux élèves, les graphiques et les symboles utilisés et tout autre renseignement jugé pertinent.

À l'exception des objectifs visés à l'intérieur des dimensions, ces renseignements sont transmis aux enseignants, par l'intermédiaire des commissions scolaires.

Le document d'information concernant la spécification de l'épreuve s'adresse donc au personnel enseignant.

3. L'APPROCHE DU MEQ

Plus le nombre d'élèves à qui s'adresse l'épreuve est grand, plus les concepteurs et les rédacteurs doivent faire des choix contraignants afin de s'assurer que l'épreuve sera valide et fidèle.

L'enseignant qui rédige une épreuve pour son groupe d'élèves a beaucoup moins de contraintes pour s'assurer que l'épreuve soit valide et fidèle. Il lui est possible d'utiliser un plus grand nombre de scénarios se rapprochant même parfois de certains objectifs terminaux; il est le seul correcteur; il peut aisément vérifier si le vocabulaire qu'il utilise est adéquat pour ses élèves, que la longueur de l'épreuve est respectable, que les conditions de passation sont uniformes pour tous et que ses élèves ont déjà rencontré le genre d'items qu'il construit.

À l'opposé, si l'épreuve est administrée à l'échelle nationale, les concepteurs rencontrent plus de contraintes que pour une épreuve à l'échelle provinciale.

Les concepteurs d'épreuves à l'échelle provinciale doivent donc mettre en place des moyens supplémentaires afin de s'assurer que l'épreuve sera aussi valide et fidèle qu'une épreuve interne.

Quelles sont donc les contraintes que rencontrent les concepteurs d'une épreuve à l'échelle provinciale?

Les contraintes que rencontrent les concepteurs d'une épreuve administrée à l'échelle provinciale.

Contraintes qui risquent d'affecter surtout la validité de l'épreuve :

Contraintes qui risquent d'affecter surtout la fidélité de l'épreuve :

Lorsqu'un enseignant rédige une épreuve qui sera administrée à ses élèves, il peut facilement et rapidement répondre adéquatement à la majorité des questions que nous venons de soulever. Il est le seul rédacteur, le seul correcteur et il connaît bien TOUS les élèves à qui s'adresse l'épreuve.

Pour que le MEQ puisse concevoir des épreuves à l'échelle provinciale valides et fidèles, il doit donc recourir à des outils supplémentaires comme la définition du domaine et la spécification de l'épreuve, et ce, avant même qu'il pense à analyser son épreuve en fonction des pistes fournies pour que les items et l'ensemble de l'épreuve soient considérés comme valides et fidèles.

De plus, il doit arrêter son choix sur une grille de correction simple et précise afin que l'ensemble des correcteurs de la province l'interprète de la même façon.

Le cadre général de l'élaboration des épreuves uniques décrit l'ensemble des opérations et des outils mis en place par le MEQ afin d'obtenir des épreuves de qualité.

Les principales étapes sont les suivantes :

1. la définition du domaine;
2. la spécification de l'épreuve;
3. la production des items;
4. la vérification de l'épreuve et la traduction;
 - sur le plan du contenu et de la docimologie,
 - sur le plan linguistique,
 - la mise à l'essai;
5. le montage de l'épreuve;
6. la composition typographique.

VALIDITÉ ET FIDÉLITÉ

Associez les qualités VALIDITÉ et FIDÉLITÉ aux principales étapes de la préparation des épreuves à l'échelle provinciale. Inscrivez le nom de la qualité vis-à-vis les étapes. Il se peut qu'une étape vise à la fois à améliorer la validité et la fidélité de l'épreuve.

Les principales étapes de la rédaction des épreuves à l'échelle provinciale sont les suivantes :

1. la définition du domaine; _____
2. la spécification de l'épreuve; _____
3. la production des items; _____
4. la vérification de l'épreuve et la traduction; _____
 - a) sur le plan du contenu et de la docimologie, _____
 - b) sur le plan linguistique, _____
 - c) la mise à l'essai; _____
5. le montage de l'épreuve; _____
6. la composition typographique. _____

4. L'APPROCHE DE L'ENSEIGNANT

Pour l'enseignant, la rédaction d'une épreuve sommative valide et fidèle comporte moins d'étapes. On peut les résumer comme suit :

1. En partant de la définition du domaine ou d'un tableau de spécifications, l'enseignant précise les objectifs et les habiletés qui seront évalués;

(Cette étape améliore la validité de votre épreuve.)

2. Il produit les items en fonction des objectifs et des habiletés qu'il a prédéterminés;

3. Il vérifie l'épreuve :

- en s'assurant de la congruence des items avec les objectifs et les habiletés,

(Cette étape améliore la validité de votre épreuve.)

- en s'assurant que les items soient bien construits,

(Cette étape améliore la validité et la fidélité de votre épreuve.)

- en examinant l'aspect linguistique,

(Cette étape améliore la fidélité de votre épreuve.)

4. Enfin, il fait le montage de l'épreuve.

5. LA RÉDACTION D'ITEMS

5.1 Les composantes de l'item

La mise en situation *Elle fournit les circonstances générales de l'item à partir de contextes.*

Les données *Les données du problème.*

La tâche à accomplir *Elle présente ce que l'élève doit faire pour que l'on puisse juger de son habileté ou de ses connaissances.*

La consigne *Elle indique à l'élève comment répondre.*

5.2 Les types d'items

Les principaux types d'items utilisés en sciences sont les suivants :

- Item à réponse choisie,
- Item de type appariement,
- Item à réponse courte,
- Item à réponse développée.

Vous trouverez les principes de rédaction propres à chaque type d'item à l'intérieur du GUIDE 2 (G-2).

Il faut se rappeler que la nature de l'objectif et l'habileté à mesurer sont plus importantes que le type d'item à utiliser. Choisissez le type d'item en fonction de l'objectif et de l'habileté.

Parfois, nous éprouvons de la difficulté à rédiger un item. Bien souvent, en changeant simplement le type d'item, nous en facilitons la rédaction.

5.3 Exemples d'items mesurant les habiletés

SCIENCES PHYSIQUES HABILITÉ _____

OBJECTIF : _____

ITEM

Mise en situation _____

Données _____

Tâche _____

Consigne _____

SOLUTION

Item à réponse choisie : les leurres et la bonne réponse.

Item à réponse construite :

réponse courte : réponse attendue et autres réponses acceptées.

réponse développée : démarche et résultat.

5.4 Les éléments essentiels d'un bon item

SYNTHÈSE DE L'ITEM

POUR UN ITEM DE QUALITÉ, ASSUREZ-VOUS :

- 1- que le choix de l'objectif est pertinent, ☐
- 2- de la congruence de l'item avec l'objectif et l'habileté mesurés, ☐
- 3- que les consignes sont claires, précises et complètes, ☐
- 4- que les élèves ont déjà une connaissance du type d'item utilisé, ☐
- 5- que le type d'item utilisé est approprié pour mesurer l'objectif au regard d'une habileté, ☐
- 6- que l'item est exempt de tout stéréotype, ☐
- 7- que le nombre de points accordés à chaque item est indiqué, ☐
- 8- que les principes de rédaction propres à chaque item sont respectés, ☐

6. LE MONTAGE DE L'ÉPREUVE

Le MEQ groupe ses items par type d'item et par thème ou par habileté. Il tient également compte de la difficulté.

Vous pouvez grouper vos items de la même façon ou choisir tout autre montage qui facilite le travail de l'élève sans causer d'ambiguïté.

SYNTHÈSE DE L'ÉPREUVE

POUR UNE ÉPREUVE VALIDE, ASSUREZ-VOUS :

1- que l'ensemble des objectifs et des habiletés que vous mesurez sont représentatifs d'une partie du programme d'études ou du programme d'études en entier, ☐

2- qu'il y a congruence entre vos items et vos objectifs et habiletés. ☐

POUR UNE ÉPREUVE FIDÈLE, ASSUREZ-VOUS MAINTENANT :

1- que la présentation est claire et lisible, ☐

2- que l'épreuve est exempte d'erreurs grammaticales, orthographiques et typographiques, ☐

3- que les directives de l'épreuve sont complètes et bien en évidence, ☐

4- que chaque item apparaît sur une même page, ☐

5- que chaque item est indépendant des autres items, ☐

6- que les critères de correction sont clairs, ☐

7- que la pondération des items représente bien l'importance des objectifs et des habiletés à l'intérieur du programme d'études, ☐

8- que la longueur de l'épreuve est appropriée par rapport au temps alloué, ☐

9- que les conditions de passation seront les mêmes pour tous. ☐

Vous trouverez, dans les pages qui suivent, les guides suivants :

GUIDE 1 : GRILLE D'ANALYSE DE L'ITEM

GUIDE 2 : LES PRINCIPES DE RÉDACTION DE CHAQUE TYPE
D'ITEM

GUIDE 3 : GUIDE DE RÉVISION DE L'ENSEMBLE DE L'ÉPREUVE

GRILLE D'ANALYSE DE L'ITEM

1. LE CHOIX DE L'OBJECTIF EST-IL PERTINENT?
(LA PERTINENCE) ☐
2. EST-CE QUE L'ITEM CORRESPOND À L'OBJECTIF
DONT IL DOIT MESURER L'ATTEINTE? (LA CONGRUENCE) ☐
3. LES CONNAISSANCES OU LES HABILÉTÉS QUE L'ITEM MESURE
SONT-ELLES SUFFISAMMENT IMPORTANTES POUR QU'ON LES MESURE? ☐
4. LA MISE EN SITUATION EST-ELLE CONCISE ET PERTINENTE? ☐
5. LA TÂCHE EST-ELLE PRÉCISE ET CLAIRE? ☐
6. LES DONNÉES SONT-ELLES ADAPTÉES AU CONTEXTE? ☐
7. LES CONSIGNES SONT-ELLES PRÉCISES, CLAIRES ET COMPLÈTES?
(LE NOMBRE DE BONNES RÉPONSES S'IL Y EN A PLUS D'UNE,
L'ENDROIT OU L'ÉLÈVE DOIT INDiquer SA RÉPONSE ET LA
FAÇON DE L'EXPRIMER, L'ENDROIT OU L'ÉLÈVE EXÉCUTE SES
CALCULS, ETC.) ☐
8. LE TYPE D'ITEM EST-IL APPROPRIÉ POUR MESURER L'OBJECTIF
RETENU? ☐
9. LE VOCABULAIRE EST-IL ADAPTÉ AUX ÉLÈVES? ☐
10. LES RÈGLES DE LA SYNTAXE SONT-ELLES RESPECTÉES? ☐
11. LES ÉLÈVES LE CONNAISSENT-ILS LE TYPE D'ITEM UTILISÉ? ☐
12. LES RENSEIGNEMENTS SUPERFLUS DONT LA PRÉSENCE PISQUE
DE COMPLIQUER INUTILEMENT LA TÂCHE DE L'ÉLÈVE SONT-ILS
ÉLIMINÉS? ☐

GRILLE D'ANALYSE DE L'ITEM
SUITE

- 13. L'ITEM EST-IL EXEMPT DE TOUTE ABERRATION SUSCEPTIBLE DE COMPROMETTRE LES RÉSULTATS? ☐
- 14. L'ITEM EST-IL EXEMPT DE TOUT STÉRÉOTYPE CULTUREL, ETHNIQUE, SOCIAL ET SEXISTE? ☐
- 15. L'ITEM EST-IL EXEMPT DE PHRASES TIRÉES DE MANUELS SCOLAIRES OU DE NOTES DE COURS? ☐
- 16. LA SOURCE DES CITATIONS EST-ELLE INDIQUÉE? ☐
- 17. LE NOMBRE DE POINTS ACCORDÉS À CHAQUE ITEM EST-IL INDIQUÉ? ☐
- 18. CHAQUE ITEM APPARAÎT-IL SUR UNE MÊME PAGE? ☐
- 19. LES PRINCIPES DE RÉDACTION PROPRES À CHAQUE TYPE D'ITEM SONT-ILS RESPECTÉS? ☐

**LES PRINCIPES DE RÉDACTION PROPRES À CHAQUE TYPE
D'ITEM SONT-ILS RESPECTÉS?**

ITEM À RÉPONSE CHOISIE

1. L'ÉNONCÉ CONTIENT TOUTES LES LIMITES ET LES CONDITIONS
NÉCESSAIRES À LA RÉOLUTION DU PROBLÈME. ☐
2. UTILISER LA PHRASE IMPÉRATIVE OU INTERROGATIVE. ☐
3. SÉPARER LA CONSIGNE DE L'ÉNONCÉ. ☐
4. PLACER DANS L'ÉNONCÉ TOUS LES MOTS QUI AUTREMENT
SE RÉPÉTERAIENT DANS CHAQUE RÉPONSE SUGGÉRÉE. ☐
5. CHAQUE RÉPONSE SUGGÉRÉE DOIT ÊTRE CONSISTANTE ET
APPROPRIÉE À L'ÉNONCÉ (À PEU PRÈS DE MÊME LONGUEUR
DE MÊME COMPLEXITÉ ET AYANT LE MÊME ATTRAIT). ☐
6. LES RÉPONSES SUGGÉRÉES DOIVENT ÊTRE CONCISES ET
SIMPLES (À MOINS QUE LA NATURE DU PROBLÈME À
RÉSOUTRE NE PRESCRIVE LE CONTRAIRE). ☐
7. AUTANT QUE POSSIBLE, LES RÉPONSES SUGGÉRÉES DOIVENT
ÊTRE PLACÉES EN COLONNE PLUTÔT QU'EN TANDEM. ☐
8. IL EST PLUS FACILE DE LIRE LES RÉPONSES QUI SONT
PLACÉES EN COLONNE PLUTÔT QU'EN TANDEM. ☐
9. ÉVITER LES NÉGATIONS AMBIGUËS. ☐
10. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
11. INDiquer LE NOMBRE DE POINTS ACCORDÉS. ☐

ITEM DE TYPE VRAI OU FAUX

1. PRÉSENTER UNE SEULE IDÉE IMPORTANTE. ☐
2. PRÉSENTER DES ITEMS MANIFESTEMENT VRAIS OU FAUX. ☐
3. PRÉSENTER DES ITEMS PLUTÔT BREFS QUE LONGS. ☐
4. ÉTOFFER SUFFISAMMENT L'ÉNONCÉ POUR QUE L'ÉLÈVE
DISPOSE DES CRITÈRES NÉCESSAIRES À LA RÉOLUTION
DU PROBLÈME. ☐
5. ÉVITER LES EXTRAITS TEXTUELS. ☐
6. UTILISER AVEC PRUDENCE LES TERMES - TOUS - TOUJOURS
JAMAIS, QUI SONT ASSOCIÉS À DES ÉNONCÉS FAUX; LES
TERMES - PARFOIS - SOUVENT - LA PLUPART, QUI SONT
ASSOCIÉS À DES ÉNONCÉS VRAIS ET LES TERMES - FRÉQUENT
NON'BREUX - UNE FOULE - PEU, QUI SONT AMBIGUS. ☐
7. IL EST SOUHAITABLE DE PLACER DANS UNE ÉPREUVE PLUS
D'ITEMS FAUX. ☐
8. ÉVITER LES NÉGATIONS AMBIGUËS. ☐
9. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
10. INDIQUER LE NOMBRE DE POINTS ACCORDÉS. ☐

ITEM DE TYPE APPARIEMENT

1. DONNER LE SCHÉMA DE BASE DE L'ASSOCIATION ET DÉCRIRE CHAQUE ENSEMBLE. ☐
2. INDIQUER S'IL Y A OU NON RÉPÉTITION DES ÉLÉMENTS DE L'ENSEMBLE-RÉPONSES. ☐
3. PLACER DANS CHAQUE ENSEMBLE DES ÉLÉMENTS HOMOGÈNES. ☐
4. NE PAS COMPTER AU TOTAL PLUS DE DIX ÉLÉMENTS DANS L'ENSEMBLE-RÉPONSES. (SIX POUR LE PRIMAIRE) ☐
5. PLACER UN PLUS GRAND NOMBRE D'ÉLÉMENTS DANS L'ENSEMBLE-RÉPONSES. ☐
6. DÉSIGNER L'ENSEMBLE-STIMULI PAR DES NOMBRES ET L'ENSEMBLE-RÉPONSES PAR DES LETTRES. ☐
7. HABITUELLEMENT L'ENSEMBLE-STIMULI EST PLACÉ À GAUCHE. ☐
8. PLACER LES ÉLÉMENTS DE L'ENSEMBLE-RÉPONSES DANS UN ORDRE QUI FACILITE LE TRAVAIL DE L'ÉLÈVE. ☐
9. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
10. INDIQUER LE NOMBRE DE POINTS ACCORDÉS. ☐

ITEM DE TYPE PHRASE À COMPLÉTER

1. ON NE DOIT ENLEVER QUE LES TERMES IMPORTANTS DE LA PHRASE. ☐
2. NE PLACER QU'UNE OU DEUX OMISSIONS PAR PHRASE À COMPLÉTER. ☐
3. PLACER LES OMISSIONS VERS LA FIN DE LA PHRASE. ☐
4. LA LONGUEUR DE L'ESPACE DOIT CORRESPONDRE À LA LONGUEUR DU MOT OMIS. ☐
5. SI LA RÉPONSE PORTE SUR DES QUANTITÉS, INDIQUER LES UNITÉS DEMANDÉES ET LE DEGRÉ DE PRÉCISION. (SI DÉSIRÉ) ☐
6. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
7. INDIQUER LE NOMBRE DE POINTS ACCORDÉS. ☐

**ITEM À RÉPONSE CONSTRUITE
(RÉPONSE COURTE)**

1. IL EST PRÉFÉRABLE DE PRÉSENTER LE PROBLÈME EN
POSANT UNE QUESTION DIRECTE PLUTÔT QU'EN
FORMULANT UNE AFFIRMATION. ☐
2. ÉVITER LES EXPRESSIONS TEXTUELLES. ☐
3. IL FAUT PRÉVOIR UN ESPACE APRÈS CHAQUE ITEM POUR
QUE L'ÉLÈVE PUISSE ÉCRIRE SA RÉPONSE IMMÉDIATEMENT
À MOINS QU'IL Y AIT UNE FEUILLE-RÉPONSES. ☐
4. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
5. INDIQUER LE NOMBRE DE POINTS ACCORDÉS. ☐

**ITEM À RÉPONSE CONSTRUITE
(RÉPONSE DÉVELOPPÉE)**

1. IL FAUT STRUCTURER, ENCADRER SUFFISAMMENT LE PROBLÈME PRÉSENTÉ DANS L'ITEM POUR QUE LA RÉPONSE SOIT ORIENTÉE VERS CE QUE L'ON SE PROPOSE DE MESURER. ☐
2. MIEUX VAUT INDiquer L'ORDRE DE GRANDEUR DE LA RÉPONSE SOUHAITÉE. ☐
3. PRÉVOIR L'ESPACE NÉCESSAIRE POUR LA SOLUTION. ☐
4. RÉDIGER DES DIRECTIVES CLAIRES ET COMPLÈTES. ☐
5. INDiquer LE NOMBRE DE POINTS ACCORDÉS ET, SI NÉCESSAIRE, LE TEMPS NORMALEMENT ALLOUÉ POUR RÉSOUDRE LE PROBLÈME. ☐
6. SI L'ENSEIGNANT LE JUGE À PROPOS, IL PEUT AVERTIR L'ÉLÈVE QU'IL SERA PÉNALISÉ S'IL INTRODUIT DANS SA RÉPONSE DES ÉLÉMENTS ÉTRANGERS AU PROBLÈME À RÉSOUDRE. ☐

GUIDE DE RÉVISION D'UNE ÉPREUVE

G-3

POUR L'ENSEMBLE DE L'ÉPREUVE

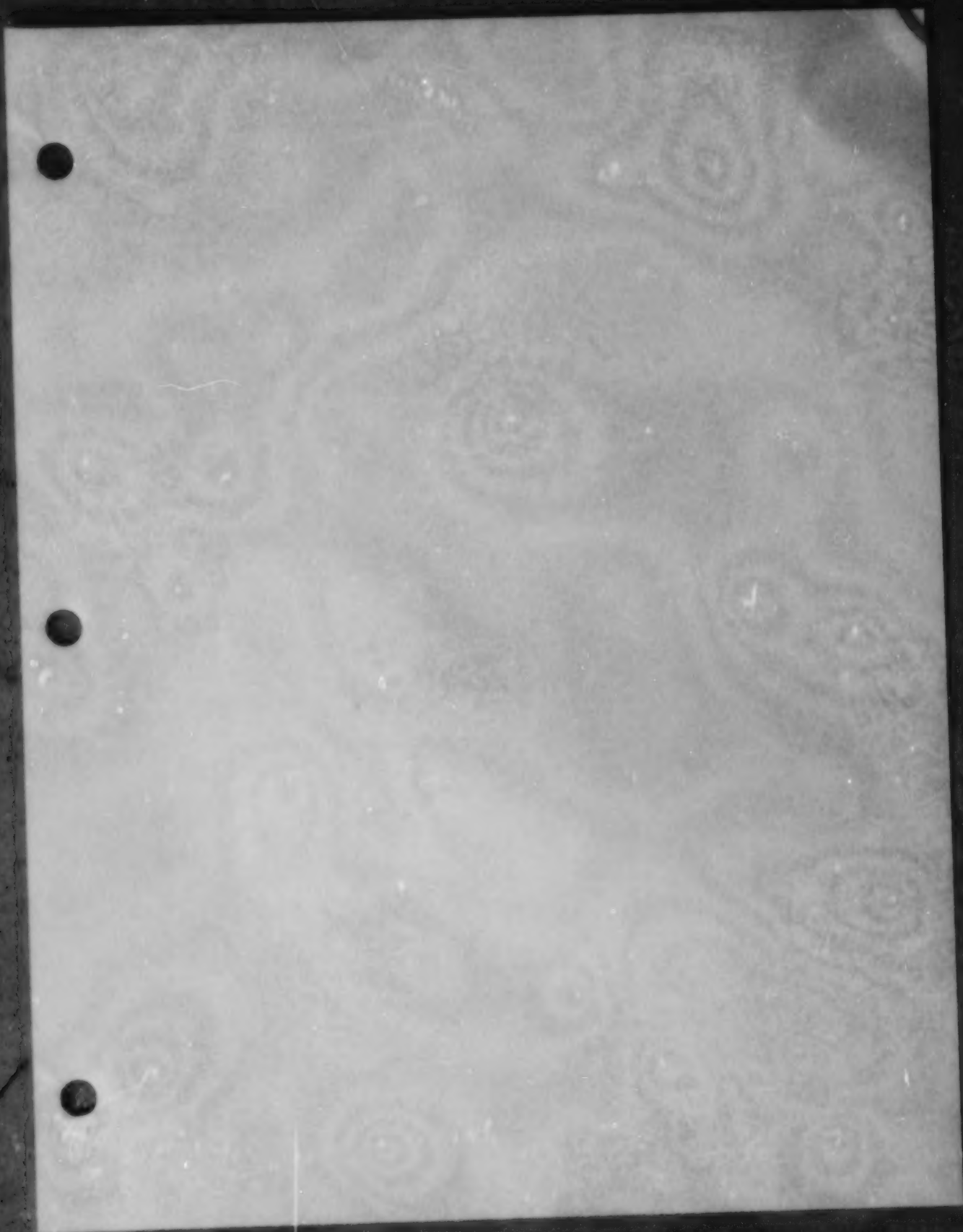
1. LE BUT DE L'ÉPREUVE EST-IL CLAIREMENT ÉTABLI? ☐
2. Y-A-IL CONGRUENCE ENTRE LE CONTENU DE L'ÉPREUVE ET LES OBJECTIFS QUE L'ON VEUT ÉVALUER? ☐
3. LE DEGRÉ DE DIFFICULTÉ DE L'ÉPREUVE EST-IL APPROPRIÉ, ÉTANT DONNÉ LE BUT DE L'ÉVALUATION ET LE NIVEAU DES ÉLÈVES? ☐
4. LE CHOIX DU TYPE D'ITEM EST-IL JUDICIEUX :
LE TYPE D'ITEM EST-IL APPROPRIÉ POUR MESURER L'OBJECTIF RETENU? ☐
5. L'AGENCEMENT DES ITEMS EST-IL PERTINENT?
(PAR TYPE D'ITEM, PAR ORDRE DE DIFFICULTÉ OU PAR ORDRE DE PRÉSENTATION DE LA MATIÈRE) ☐
6. LA LONGUEUR DE L'ÉPREUVE EST-ELLE APPROPRIÉE PAR RAPPORT AU TEMPS ALLOUÉ, AU NOMBRE D'ITEMS ET AU NIVEAU DES ÉLÈVES? ☐
7. LES DIRECTIVES SONT-ELLES CLAIRES, COMPLÈTES, PRÉCISES ET BIEN EN ÉVIDENCE?
(LE TEMPS ALLOUÉ, L'ENDROIT OÙ L'ÉLÈVE DOIT INSCRIRE SA RÉPONSE ET LA FAÇON DE L'EXPRIMER, LE NOMBRE DE BONNE RÉPONSES S'IL Y EN A PLUS D'UNE, L'ENDROIT OÙ L'ÉLÈVE EXÉCUTE SES CALCULS, LA VALEUR DE L'ÉPREUVE, ETC.) ☐
8. LES CONSIGNES POUR RÉPONDRE AUX ITEMS SONT-ELLES CLAIRES ET RÉDIGÉES AVEC UN VOCABULAIRE ACCESSIBLE AUX ÉLÈVES? ☐
9. LES TEXTES UTILISÉS SONT-ILS ADAPTÉS AUX INTÉRÊTS ET AU NIVEAU DES ÉLÈVES? ☐

**POUR L'ENSEMBLE DE L'ÉPREUVE
SUITE**

- 10. LE NIVEAU DE LANGUE (VOCABULAIRE ET SYNTAXE)
EST-IL ADAPTÉ AUX ÉLÈVES? ☐
- 11. CHAQUE ITEM EST-IL INDÉPENDANT DES AUTRES ITEMS? ☐
- 12. L'ÉPREUVE EST-ELLE EXEMPTÉ D'ERREURS GRAMMATICALES,
ORTHOGRAPHIQUES ET TYPOGRAPHIQUES? ☐
- 13. LES ILLUSTRATIONS SONT-ELLES CLAIRES? ☐
- 14. LES CARACTÈRES D'IMPRIMERIE SONT-ILS LISIBLES? ☐

POUR LA CLÉ DE CORRECTION

1. LA RÉPONSE EST-ELLE CLAIREMENT INDIQUÉE POUR
CHAQUE ITEM? ☐
2. Y A-T-IL D'AUTRES RÉPONSES ACCEPTABLES POUR
CHAQUE ITEM? ☐
3. LES AUTRES RÉPONSES ACCEPTABLES SONT-ELLES INDIQUÉES? ☐
4. LE NOMBRE DE POINTS ACCORDÉS À CHAQUE ITEM OU ASPECT
DE L'ITEM EST-IL INDIQUÉ? ☐
5. LES CRITÈRES DE CORRECTION POUR CHAQUE ITEM À RÉPONSE
CONSTRuite SONT-ILS INDIQUÉS? ☐
6. LES CRITÈRES DE CORRECTION SONT-ILS CLAIRS? ☐



ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 5
La correction de l'épreuve
GUIDE DE L'ANIMATEUR

ACTIVITÉ 5

La correction de l'épreuve

1. *L'évaluation des habiletés reliées à la démarche expérimentale par l'intermédiaire du rapport de laboratoire*
2. *L'évaluation de l'habileté à résoudre des problèmes par l'intermédiaire de l'épreuve «papier-crayon»*
3. *La correction de l'enseignant*

1. L'ÉVALUATION DES HABILITÉS RELIÉES À LA DÉMARCHE EXPÉRIMENTALE PAR L'INTERMÉDIAIRE DU RAPPORT DE LABORATOIRE

Il est beaucoup plus facile de vérifier si un élève connaît telle loi de la physique que de juger de la maîtrise des habiletés reliées à la démarche expérimentale. Voyons quelques raisons du fait qu'il en est ainsi.

Dans le cas de la connaissance d'une loi de physique, l'évaluation s'applique principalement à un contenu, la plupart du temps très particulier et délimité par un programme d'études.

Dans le cas de la démarche expérimentale, l'évaluation s'applique à des habiletés qui sont, autrement dit, des construits au même titre que l'intelligence, la créativité, etc. L'évaluation de construits porte sur des objets aux contours plus flous, moins directement observables et au sujet desquels le consensus est plus difficile à faire.

Si on ne peut observer directement un construit ou une habileté, on peut par ailleurs porter un jugement à partir du produit que la mise en oeuvre de cette habileté a permis de créer. Par exemple, un rapport d'analyse donnera un bon indice de la capacité d'analyser de son auteur. Dans ce cas, le jugement évaluatif s'appuie sur divers critères comme la présence ou l'absence de certains éléments ou leur teneur par rapport à certaines caractéristiques désirables. Dans le cas de la mesure des habiletés reliées à la démarche expérimentale, le rapport de laboratoire est un excellent produit pour nous permettre d'évaluer ces habiletés.

Certes, il est difficile d'évaluer un ensemble d'habiletés aussi complexes que celles qui sont reliées à la démarche expérimentale, principalement dans le cadre d'une épreuve à l'échelle provinciale où les contraintes sont nombreuses.

Il importe donc de faire émerger, à travers les difficultés, la procédure d'évaluation méthodique et rigoureuse qui permettra d'évaluer de la façon la plus juste possible.

Pour évaluer les habiletés reliées à la démarche expérimentale, au moment d'établir les critères d'évaluation mesurables et observables du rapport de laboratoire, il ne faut pas oublier que l'expérimentation nous apporte beaucoup d'éléments de réponse.

Voici donc où est rendue la réflexion concernant l'évaluation des habiletés reliées à la démarche expérimentale par l'intermédiaire du rapport de laboratoire.

LA GRILLE DE CORRECTION DU MINISTÈRE DE L'ÉDUCATION

GRILLE DE NOTATION

ÉTAPE	CRITÈRES	ÉCHELLE				
HYPOTHÈSE 5 %	Proposition					
	. PERTINENCE	5	4	3	1	0
PROTOCOLE 40 %	Choix des variables					
	. PERTINENCE ET NOMBRE SUFFISANT	20	15	12	5	0
	Matériel et manipulation					
	. USAGE ET PERTINENCE	20	15	12	5	0
RÉUNION DES DONNÉES 25 %	Données					
	. CLARTÉ DE LA PRÉSENTATION	5	4	3	1	0
	. PERTINENCE	10	8	6	3	0
	. PRÉCISION	10	8	6	3	0
ORGANISATION, INTERPRÉTATION ET CONCLUSION 30 %	Interprétation					
	. LOGIQUE DES INFÉRENCES	10	8	6	3	0
	Conclusion					
	. JUSTIFICATION	20	15	12	5	0
		100	75	60	25	0

NOTE. Le modèle et l'importance relative des critères peuvent changer légèrement à cause de la nature de certains problèmes.

INTERPRÉTATION DE LA GRILLE DE NOTATION

NOTE	SIGNIFICATION
20-10-05	<i>Tous les éléments sont présents et acceptables. La qualité du travail est excellente à l'étape considérée.</i>
15-08-04	<i>La majorité des éléments sont présents et acceptables. La qualité est très bonne ou supérieure à la moyenne.</i>
12-06-03	<i>Certains éléments sont acceptables. La qualité est suffisante. (Seuil de réussite)</i>
05-03-01	<i>Le nombre d'éléments acceptables est insuffisant.</i>
00-00-00	<i>Les éléments présents sont inacceptables ou il n'y a aucun élément.</i>

2. L'ÉVALUATION DE L'HABILETÉ À RÉSCUDRE DES PROBLÈMES PAR L'INTERMÉDIAIRE DE L'ÉPREUVE «PAPIER-CRAYON»

L'évaluation juste de l'habileté à résoudre des problèmes à l'intérieur d'une épreuve «papier-crayon» est sans doute une des plus difficiles à réaliser surtout si l'on tient compte de la démarche de l'élève.

L'habileté à résoudre des problèmes est entière, c'est-à-dire qu'on peut difficilement affirmer que l'élève a démontré l'habileté par rapport à un item s'il a résolu partiellement le problème.

Rappelons que la démarche de l'élève, c'est tout ce qu'il fait et pense pendant qu'il résout le problème et ce que l'on connaît de sa démarche, ce sont les traces qu'il veut bien nous laisser.

Dans le contexte d'une évaluation formative, l'évaluation de la résolution de problèmes cause beaucoup moins de difficultés, car l'objectif premier de l'évaluation formative est de favoriser l'apprentissage de l'élève. Tous les moyens sont bons pour aider l'élève dans son apprentissage; que l'on parle du travail en équipe, de grilles d'observation utilisées par l'élève, un camarade ou l'enseignant, d'une échelle d'attitudes, d'items à réponse construite ou à réponse choisie, d'interviews, etc. La note totale n'est pas d'une grande utilité, c'est la qualité de la rétroaction à l'élève qui est essentielle.

Dans une évaluation sommative, comme le but premier est de reconnaître des acquis ou de certifier des apprentissages, les scénarios doivent être plus rigides. La note totale devient importante, car elle est la représentation de la performance de l'élève; cette note est habituellement consignée à l'intérieur d'un formulaire de consignation des notes.

Dans ce contexte, il devient essentiel de s'assurer de la justesse de la note de l'élève; ce qui rend la chose difficile, c'est qu'une grande partie de la démarche de résolution du problème se passe dans la tête de l'élève.

LA GRILLE DE CORRECTION DU MEQ

Cette grille est conçue pour les items à réponse construite développée. En ce qui concerne les questions à réponse choisie ou à réponse construite courte, la correction est dichotomique, 4 points ou 0.

0	2	3	4
---	---	---	---

Règles générales :

- 4 Démarche complète acceptable.
Résultat exact.
- 3 Démarche complète acceptable.
Résultat inexact à cause d'une erreur de calcul ou de transcription.
- 2 Démarche partielle acceptable montrant que l'élève a suivi la démarche scientifique et a atteint l'objectif et l'habileté visés.
(L'attribution de cette note est occasionnelle.)
- 0 Démarche absente ou erronée.

L'application intégrale ou partielle de ce modèle est fonction de la nature de l'habileté que l'item permet de mesurer.

Règles particulières :

Pour chaque item, on peut décrire d'une façon particulière chacun des échelons.

Ces façons de faire du MEQ ne sont pas des modèles.

Ce sont tout simplement des façons justes d'évaluer certaines habiletés en tenant compte des contraintes qu'occasionnent les épreuves à l'échelle provinciale.

Quelles sont ces contraintes?

- le nombre important de correcteurs,
- la correction n'est pas contrôlée, c'est-à-dire qu'on ne met en place aucun moyen de vérifier si tous les correcteurs de la province corrigent de la même façon et si chaque correcteur corrige toutes ses copies de la même façon,
- une grille de correction simple et précise afin que l'ensemble des correcteurs de la province l'interprètent de la même façon,
- le MEQ n'a aucun contrôle sur l'élève qui répond au hasard ou qui triche; le contrôle appartient aux autorités locales en ce qui concerne l'élève qui triche (ce genre de contrôle ne peut se faire au moment de la correction sans pénaliser tous les élèves de la province et depuis longtemps on a prouvé que cette façon de faire cause plus d'erreurs de mesure qu'elle n'en règle).

Pourquoi la correction du MEQ est-elle juste pour l'élève?

Cette question nous ramène à une des qualités fondamentales de la mesure : la fidélité.

La qualité de fidélité requiert que les résultats des élèves au regard d'une épreuve soient stables. Et l'un des aspects importants à considérer pour une mesure stable, c'est la correction. Si deux correcteurs attribuent à un élève deux résultats légèrement différents, il devient difficile de parler de mesure stable. Il faut donc penser ici, pour parler de mesure fidèle (stable), que tous les correcteurs de la province attribuent le même résultat pour des copies d'épreuves semblables.

Tout un défi à relever!

C'est pourtant à ce chapitre que tout doit se jouer. Plus les correcteurs sont nombreux, plus la grille de correction doit être simple et plus elle doit comporter un nombre restreint d'échelons définis d'une façon très précise.

Une grille de correction à neuf échelons, par exemple, pourrait s'avérer un instrument de grande qualité pour un enseignant qui aurait défini très précisément chacun des échelons. Un instrument de grande qualité parce que l'enseignant, connaissant très bien chacun des échelons, pourrait corriger de façon juste chacune des copies de ses élèves. Il pourrait atteindre une discrimination beaucoup plus importante que s'il utilisait une grille du MEQ. Par contre, cette grille à neuf échelons pourrait déstabiliser les résultats des élèves si elle était utilisée par des centaines de correcteurs. Non pas parce que la grille n'est pas de qualité, mais avec un grand nombre d'utilisateurs, la probabilité que certains interprètent les échelons d'une façon différente est plus grande. Autrement dit, la qualité de la grille n'est pas une condition suffisante pour s'assurer d'obtenir des mesures fidèles. Afin d'obtenir des mesures stables lorsque plusieurs correcteurs sont en cause, il faut trouver une échelle facile à utiliser et dont le nombre d'échelons et la qualité descriptive assurent une mesure valide et fidèle.

L'habileté du correcteur à utiliser une grille de correction est également une condition préalable à l'obtention de mesures stables.

Dans ce sens, les grilles de correction du MEQ s'avèrent un compromis équitable afin de s'assurer de mesures stables. Plus les enseignants se familiariseront à l'utilisation de ces grilles, plus les mesures seront de qualité.

Il est à noter que ces grilles de correction peuvent être améliorées ou modifiées par le Ministère.

Pourquoi tenir compte des contraintes qu'occasionnent les épreuves à l'échelle provinciale au moment de la correction?

Parce que ces contraintes peuvent diminuer la fidélité de l'épreuve. Cette qualité est primordiale car des décisions relatives à la sanction des études sont en jeu.

Un petit exemple pour les fervents des statistiques et pour les fanatiques des théories de la mesure...

On arrive à estimer, pour chaque élève, un intervalle dans lequel on retrouve son score vrai.

Vous vous souvenez de l'équation : $V = X \pm e$ qui signifie que le score vrai de l'élève (V) égale son score observé (X) plus ou moins un facteur d'erreur (e).

Voici donc une façon d'interpréter le coefficient de fidélité¹.

Supposons une moyenne $\bar{X} = 60$, et un écart-type $S_x = 5$, pour une épreuve dont le coefficient de fidélité $r_{xx} = 0,70$.

L'erreur de mesure $S_e = S_x \sqrt{1 - r_{xx}}$

$$S_e = 5 \sqrt{1 - 0,70} = 2,74$$

1. Adapté de J.-J. Bernier, **Théorie des tests**, Principes et techniques de base, deuxième édition, Gaétan Morin éditeur, 1985, p. 136.

Pour un élève ayant comme score observé $X = 57$ à un niveau de confiance égal à 0,95 (ZS_e étant le score Z de l'erreur de mesure), on obtient :

$$X - ZS_e < V < X + ZS_e$$

$$57 - (1,96) (2,74) < V < 57 + (1,96) (2,74)$$

$$51,63 < V < 62,37$$

Il y a donc 95 % des chances que l'élève ayant un score observé $X = 57$ ait un score vrai entre 51,63 et 62,37.

Pour ce même élève, à un niveau de confiance de 0,99, on obtiendrait l'intervalle suivant :

$$X - ZS_e < V < X + ZS_e$$

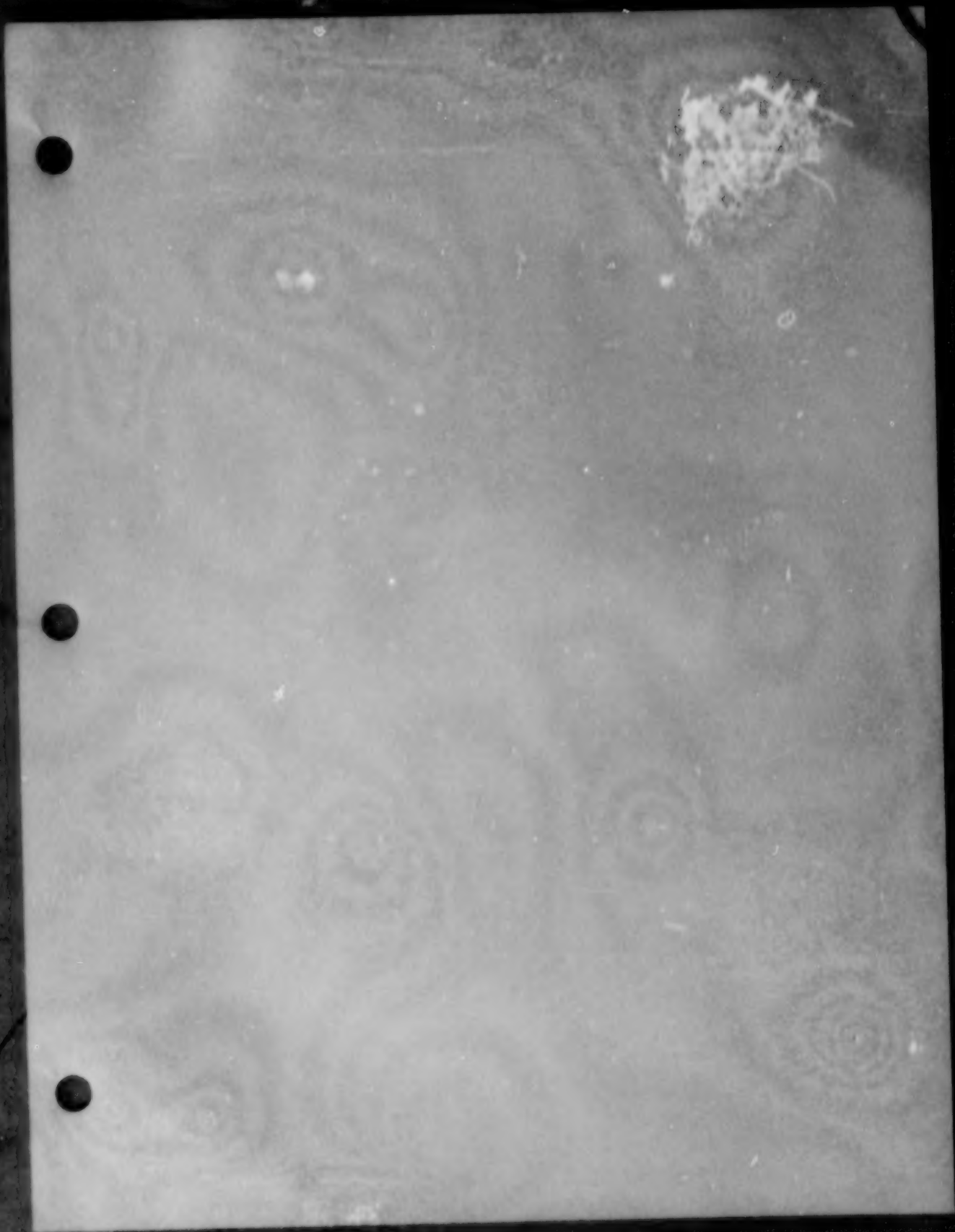
$$57 - (2,56) (2,74) < V < 57 + (2,56) (2,74)$$

$$49,99 < V < 64,01$$

Il y a donc 99 % des chances que l'élève ayant un score observé $X = 57$ ait un score vrai entre 49,99 et 64,01.

L'erreur de mesure est un indice de la variabilité des scores observés autour du score vrai. C'est une estimation du degré de modification des scores attendus au moment d'une réadministration éventuelle du test. On l'utilise comme un écart-type, celui de la distribution des erreurs. **Plus l'erreur de mesure est grande, plus l'intervalle est grand.**

À l'aide de cet exemple, on peut facilement imaginer les répercussions que pourrait avoir une épreuve à faible niveau de fidélité sur les décisions prises relativement à la sanction des études de plusieurs élèves.



Dans ce contexte, la qualité de fidélité d'une épreuve devient extrêmement importante. Le MEQ est donc obligé de tenir compte des contraintes qu'occasionne une épreuve à l'échelle provinciale s'il veut s'assurer de la fidélité d'une épreuve.

Est-ce que ces contraintes m'affectent comme enseignant?

Non, ces contraintes ne vous affectent pas, car vous êtes habituellement le seul correcteur. Il ne vous reste qu'à vous assurer que vous corrigez la dernière copie de l'épreuve de la même façon que la première. C'est déjà tout un travail! Plusieurs recherches l'ont démontré.

Concernant l'élève qui triche ou qui répond au hasard, comme vous êtes celui qui, parmi tous les agents du système d'éducation, connaissez le mieux l'élève en question, vous êtes donc la meilleure personne pour prendre une décision par rapport à cet élève. Autrement dit, vous êtes la personne la plus près de l'élève et c'est donc vous qui prendrez la décision la plus juste.

3. LA CORRECTION DE L'ENSEIGNANT

Pour corriger vos épreuves, vous pouvez appliquer les grilles du MEQ et devenir d'habiles utilisateurs et ainsi contribuer à l'obtention de mesures stables ou fidèles au niveau provincial.

Vous pouvez aussi utiliser toute autre grille de correction, car vous avez beaucoup moins de contraintes que le MEQ.

Mais vous aussi vous devez vous assurer d'obtenir des mesures stables ou fidèles.

Comment?

- *en ayant des critères de correction observables et mesurables,*
- *en ayant des échelons pertinents qui mesurent bien ce que vous voulez mesurer,*
- *en vous assurant que chacun de ces échelons soit bien décrit, afin qu'il n'y ait pas d'ambiguïté entre les différents échelons,*
- *en vous assurant que chacun de ces échelons soit indépendant des autres échelons,*
- *en vous assurant que vous êtes habile à utiliser l'instrument de votre choix.*



ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 5

La correction de l'épreuve

GUIDE DU PARTICIPANT

ACTIVITÉ 5

La correction de l'épreuve

1. *L'évaluation des habiletés reliées à la démarche expérimentale par l'intermédiaire du rapport de laboratoire*
2. *L'évaluation de l'habileté à résoudre des problèmes par l'intermédiaire de l'épreuve «papier-crayon»*
3. *La correction de l'enseignant*

1. L'ÉVALUATION DES HABILITÉS RELIÉES À LA DÉMARCHE
EXPÉRIMENTALE PAR L'INTERMÉDIAIRE DU RAPPORT DE
LABORATOIRE

Il est beaucoup plus facile de vérifier si un élève connaît telle loi de la physique que de juger de la maîtrise des habiletés reliées à la démarche expérimentale. Voyons quelques raisons du fait qu'il en est ainsi.

Dans le cas de la connaissance d'une loi de physique, l'évaluation s'applique principalement à un contenu, la plupart du temps très particulier et délimité par un programme d'études.

Dans le cas de la démarche expérimentale, l'évaluation s'applique à des habiletés qui sont, autrement dit, des construits au même titre que l'intelligence, la créativité, etc. L'évaluation de construits porte sur des objets aux contours plus flous, moins directement observables et au sujet desquels le consensus est plus difficile à faire.

Si on ne peut observer directement un construit ou une habileté, on peut par ailleurs porter un jugement à partir du produit que la mise en oeuvre de cette habileté a permis de créer. Par exemple, un rapport d'analyse donnera un bon indice de la capacité d'analyser de son auteur. Dans ce cas, le jugement évaluatif s'appuie sur divers critères comme la présence ou l'absence de certains éléments ou leur teneur par rapport à certaines caractéristiques désirables. Dans le cas de la mesure des habiletés reliées à la démarche expérimentale, le rapport de laboratoire est un excellent produit pour nous permettre d'évaluer ces habiletés.

Certes, il est difficile d'évaluer un ensemble d'habiletés aussi complexes que celles qui sont reliées à la démarche expérimentale, principalement dans le cadre d'une épreuve à l'échelle provinciale où les contraintes sont nombreuses.

Il importe donc de faire émerger, à travers les difficultés, la procédure d'évaluation méthodique et rigoureuse qui permettra d'évaluer de la façon la plus juste possible.

Pour évaluer les habiletés reliées à la démarche expérimentale, au moment d'établir les critères d'évaluation mesurables et observables du rapport de laboratoire, il ne faut pas oublier que l'expérimentation nous apporte beaucoup d'éléments de réponse.

Voici donc où est rendue la réflexion concernant l'évaluation des habiletés reliées à la démarche expérimentale par l'intermédiaire du rapport de laboratoire.

LA GRILLE DE CORRECTION DU MINISTÈRE DE L'ÉDUCATION

GRILLE DE NOTATION

ÉTAPE	CRITÈRES	ÉCHELLE				
HYPOTHÈSE 5 %	Proposition PERTINENCE	5	4	3	1	0
PROTOCOLE 40 %	Choix des variables PERTINENCE ET NOMBRE SUFFISANT	20	15	12	5	0
	Matériel et manipulation USAGE ET PERTINENCE	20	15	12	5	0
RÉUNION DES DONNÉES 25 %	Données CLARTÉ DE LA PRÉSENTATION	5	4	3	1	0
	PERTINENCE	10	8	6	3	0
	PRÉCISION	10	8	6	3	0
ORGANISATION, INTERPRÉTATION ET CONCLUSION 30 %	Interprétation LOGIQUE DES INFÉRENCES	10	8	6	3	0
	Conclusion JUSTIFICATION	20	15	12	5	0
		100	75	60	25	0

NOTE: Le modèle et l'importance relative des critères peuvent changer légèrement à cause de la nature de certains problèmes.

INTERPRÉTATION DE LA GRILLE DE NOTATION

NOTE	SIGNIFICATION
20-10-05	Tous les éléments sont présents et acceptables. La qualité du travail est excellente à l'étape considérée.
15-08-04	La majorité des éléments sont présents et acceptables. La qualité est très bonne ou supérieure à la moyenne.
12-06-03	Certains éléments sont acceptables. La qualité est suffisante. (Seuil de réussite)
05-03-01	Le nombre d'éléments acceptables est insuffisant.
00-00-00	Les éléments présents sont inacceptables ou il n'y a aucun élément.

2. L'ÉVALUATION DE L'HABILITÉ À RÉSOUDRE DES PROBLÈMES PAR L'INTERMÉDIAIRE DE L'ÉPREUVE «PAPIER-CRAYON»

L'évaluation juste de l'habileté à résoudre des problèmes à l'intérieur d'une épreuve «papier-crayon» est sans doute une des plus difficiles à réaliser surtout si l'on tient compte de la démarche de l'élève.

L'habileté à résoudre des problèmes est entière, c'est-à-dire qu'on peut difficilement affirmer que l'élève a démontré l'habileté par rapport à un item s'il a résolu partiellement le problème.

Rappelons que la démarche de l'élève, c'est tout ce qu'il fait et pense pendant qu'il résout le problème et ce que l'on connaît de sa démarche, ce sont les traces qu'il veut bien nous laisser.

Dans le contexte d'une évaluation formative, l'évaluation de la résolution de problèmes cause beaucoup moins de difficultés, car l'objectif premier de l'évaluation formative est de favoriser l'apprentissage de l'élève. Tous les moyens sont bons pour aider l'élève dans son apprentissage; que l'on parle du travail en équipe, de grilles d'observation utilisées par l'élève, un camarade ou l'enseignant, d'une échelle d'attitudes, d'items à réponse construite ou à réponse choisie, d'interviews, etc. La note totale n'est pas d'une grande utilité, c'est la qualité de la rétroaction à l'élève qui est essentielle.

Dans une évaluation sommative, comme le but premier est de reconnaître des acquis ou de certifier des apprentissages, les scénarios doivent être plus rigides. La note totale devient importante, car elle est la représentation de la performance de l'élève; cette note est habituellement consignée à l'intérieur d'un

formulaire de consignation des notes.

Dans ce contexte, il devient essentiel de s'assurer de la justesse de la note de l'élève; ce qui rend la chose difficile, c'est qu'une grande partie de la démarche de résolution du problème se passe dans la tête de l'élève.

LA GRILLE DE CORRECTION DU MEQ

Cette grille est conçue pour les items à réponse construite développée. En ce qui concerne les questions à réponse choisie ou à réponse construite courte, la correction est dichotomique, 4 points ou 0.

0	2	3	4
---	---	---	---

Règles générales :

- 4 Démarche complète acceptable.
Résultat exact.
- 3 Démarche complète acceptable.
Résultat inexact à cause d'une erreur de calcul ou de transcription.
- 2 Démarche partielle acceptable montrant que l'élève a suivi la démarche scientifique et a atteint l'objectif et l'habileté visés.
(L'attribution de cette note est occasionnelle.)
- 0 Démarche absente ou erronée.

L'application intégrale ou partielle de ce modèle est fonction de la nature de l'habileté que l'item permet de mesurer.

Règles particulières :

Pour chaque item, on peut décrire d'une façon particulière chacun des échelons.

Ces façons de faire du MEQ ne sont pas des modèles.

Ce sont tout simplement des façons justes d'évaluer certaines habiletés en tenant compte des contraintes qu'occasionnent les épreuves à l'échelle provinciale.

Quelles sont ces contraintes?

- le nombre important de correcteurs,
- la correction n'est pas contrôlée, c'est-à-dire qu'on ne met en place aucun moyen de vérifier si tous les correcteurs de la province corrigent de la même façon et si chaque correcteur corrige toutes ses copies de la même façon,
- une grille de correction simple et précise afin que l'ensemble des correcteurs de la province l'interprètent de la même façon,
- le MEQ n'a aucun contrôle sur l'élève qui répond au hasard ou qui triche; le contrôle appartient aux autorités locales en ce qui concerne l'élève qui triche (ce genre de contrôle ne peut se faire au moment de la correction sans pénaliser tous les élèves de la province et depuis longtemps on a prouvé que cette façon de faire cause plus d'erreurs de mesure qu'elle n'en règle).

Pourquoi la correction du MEQ est-elle juste pour l'élève?

Cette question nous ramène à une des qualités fondamentales de la mesure : la fidélité.

La qualité de fidélité requiert que les résultats des élèves au regard d'une épreuve soient stables. Et l'un des aspects importants à considérer pour une mesure stable, c'est la correction. Si deux correcteurs attribuent à un élève deux résultats légèrement différents, il devient difficile de parler de mesure stable. Il faut donc penser ici, pour parler de mesure fidèle (stable), que tous les correcteurs de la province attribuent le même résultat pour des copies d'épreuves semblables.

Tout un défi à relever!

C'est pourtant à ce chapitre que tout doit se jouer. Plus les correcteurs sont nombreux, plus la grille de correction doit être simple et plus elle doit comporter un nombre restreint d'échelons définis d'une façon très précise.

Une grille de correction à neuf échelons, par exemple, pourrait s'avérer un instrument de grande qualité pour un enseignant qui aurait défini très précisément chacun des échelons. Un instrument de grande qualité parce que l'enseignant, connaissant très bien chacun des échelons, pourrait corriger de façon juste chacune des copies de ses élèves. Il pourrait atteindre une discrimination beaucoup plus importante que s'il utilisait une grille du MEQ. Par contre, cette grille à neuf échelons pourrait déstabiliser les résultats des élèves si elle était utilisée par des centaines de correcteurs. Non pas parce que la grille n'est pas de qualité, mais avec un grand nombre d'utilisateurs, la probabilité que certains interprètent les échelons d'une façon différente est plus grande. Autrement dit, la qualité de la grille n'est pas une condition suffisante pour s'assurer d'obtenir des mesures fidèles. Afin d'obtenir des mesures stables lorsque plusieurs correcteurs sont en cause, il faut trouver une échelle facile à utiliser et dont le nombre d'échelons et la qualité descriptive assurent une mesure valide et fidèle.

L'habileté du correcteur à utiliser une grille de correction est également une condition préalable à l'obtention de mesures stables.

Dans ce sens, les grilles de correction du MEQ s'avèrent un compromis équitable afin de s'assurer de mesures stables. Plus les enseignants se familiariseront à l'utilisation de ces grilles, plus les mesures seront de qualité.

Il est à noter que ces grilles de correction peuvent être améliorées ou modifiées par le Ministère.

Pourquoi tenir compte des contraintes qu'occasionnent les épreuves à l'échelle provinciale au moment de la correction?

Parce que ces contraintes peuvent diminuer la fidélité de l'épreuve. Cette qualité est primordiale car des décisions relatives à la sanction des études sont en jeu.

Un petit exemple pour les fervents des statistiques et pour les fanatiques des théories de la mesure...

On arrive à estimer, pour chaque élève, un intervalle dans lequel on retrouve son score vrai.

Vous vous souvenez de l'équation : $V = X \pm e$ qui signifie que le score vrai de l'élève (V) égale son score observé (X) plus ou moins un facteur d'erreur (e).

Voici donc une façon d'interpréter le coefficient de fidélité.¹

Supposons une moyenne $\bar{X} = 60$, et un écart-type $S_x = 5$, pour une épreuve dont le coefficient de fidélité $r_{xx} = 0,70$.

L'erreur de mesure $S_e = S_x \sqrt{1 - r_{xx}}$

$$S_e = 5 \sqrt{1 - 0,70} = 2,74$$

1. Adapté de J.-J. Bernier, **Théorie des tests**, Principes et techniques de base, deuxième édition, Gaëtan Morin éditeur, 1985, p. 136.

Pour un élève ayant comme score observé $X = 57$ à un niveau de confiance égal à 0,95 (ZS_e étant le score Z de l'erreur de mesure), on obtient :

$$X - ZS_e < V < X + ZS_e$$

$$57 - (1,96) (2,74) < V < 57 + (1,96) (2,74)$$

$$51,63 < V < 62,37$$

Il y a donc 95 % des chances que l'élève ayant un score observé $X = 57$ ait un score vrai entre 51,63 et 62,37.

Pour ce même élève, à un niveau de confiance de 0,99, on obtiendrait l'intervalle suivant :

$$X - ZS_e < V < X + ZS_e$$

$$57 - (2,56) (2,74) < V < 57 + (2,56) (2,74)$$

$$49,99 < V < 64,01$$

Il y a donc 99 % des chances que l'élève ayant un score observé $X = 57$ ait un score vrai entre 49,99 et 64,01.

L'erreur de mesure est un indice de la variabilité des scores observés autour du score vrai. C'est une estimation du degré de modification des scores attendus au moment d'une réadministration éventuelle du test. On l'utilise comme un écart-type, celui de la distribution des erreurs. **Plus l'erreur de mesure est grande, plus l'intervalle est grand.**

À l'aide de cet exemple, on peut facilement imaginer les répercussions que pourrait avoir une épreuve à faible niveau de fidélité sur les décisions prises relativement à la sanction des études de plusieurs élèves.

Dans ce contexte, la qualité de fidélité d'une épreuve devient extrêmement importante. Le MEQ est donc obligé de tenir compte des contraintes qu'occasionne une épreuve à l'échelle provinciale s'il veut s'assurer de la fidélité d'une épreuve.

Est-ce que ces contraintes m'affectent comme enseignant?

Non, ces contraintes ne vous affectent pas, car vous êtes habituellement le seul correcteur. Il ne vous reste qu'à vous assurer que vous corrigez la dernière copie de l'épreuve de la même façon que la première. C'est déjà tout un travail! Plusieurs recherches l'ont démontré.

Concernant l'élève qui triche ou qui répond au hasard, comme vous êtes celui qui, parmi tous les agents du système d'éducation, connaissez le mieux l'élève en question, vous êtes donc la meilleure personne pour prendre une décision par rapport à cet élève. Autrement dit, vous êtes la personne la plus près de l'élève et c'est donc vous qui prendrez la décision la plus juste.

3. LA CORRECTION DE L'ENSEIGNANT

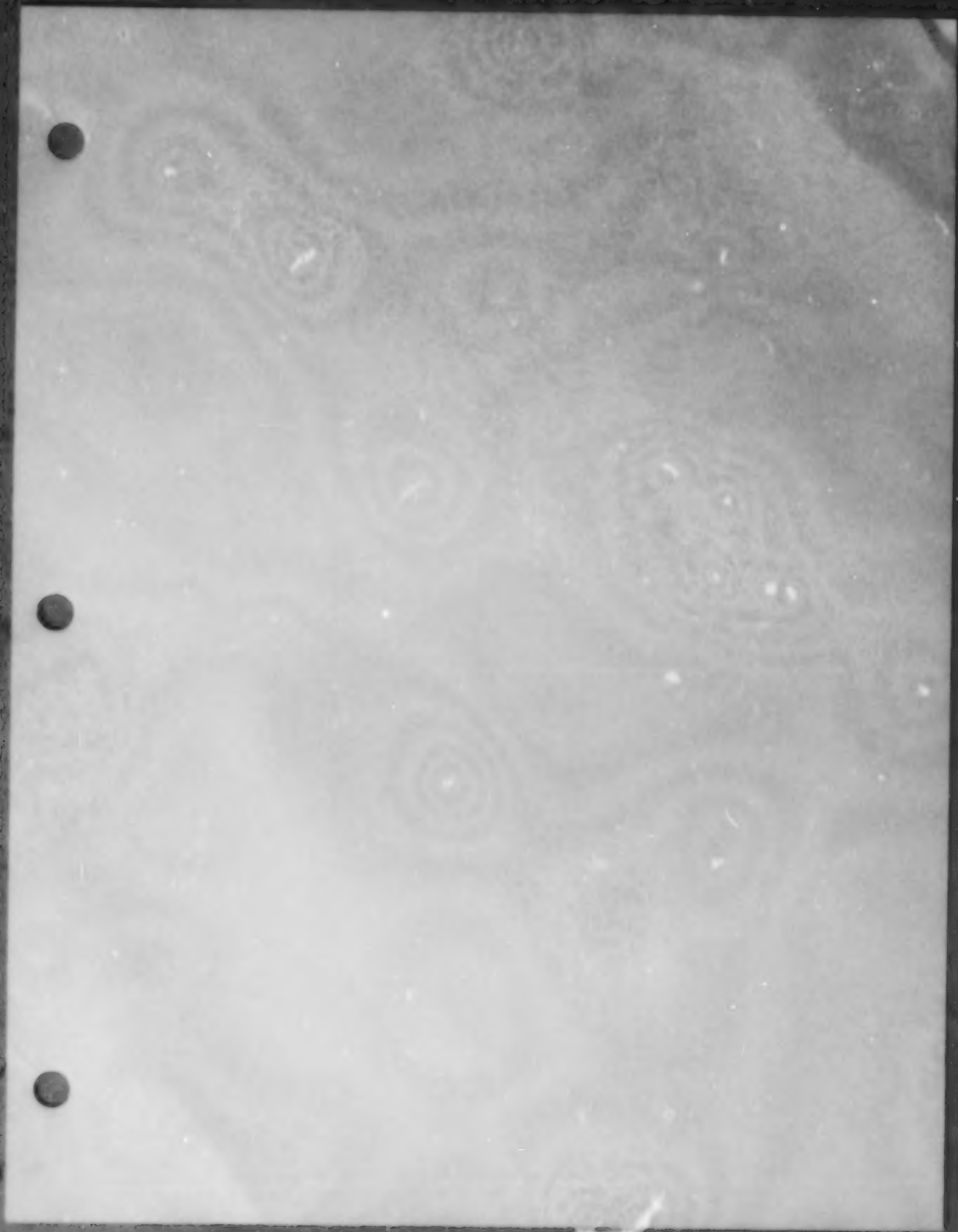
Pour corriger vos épreuves, vous pouvez appliquer les grilles du MEQ et devenir d'habiles utilisateurs et ainsi contribuer à l'obtention de mesures stables ou fidèles au niveau provincial.

Vous pouvez aussi utiliser toute autre grille de correction, car vous avez beaucoup moins de contraintes que le MEQ.

Mais vous aussi vous devez vous assurer d'obtenir des mesures stables ou fidèles.

Comment?

- *en ayant des critères de correction observables et mesurables,*
- *en ayant des échelons pertinents qui mesurent bien ce que vous voulez mesurer,*
- *en vous assurant que chacun de ces échelons soit bien décrit, afin qu'il n'y ait pas d'ambiguïté entre les différents échelons,*
- *en vous assurant que chacun de ces échelons soit indépendant des autres échelons,*
- *en vous assurant que vous êtes habile à utiliser l'instrument de votre choix.*



ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 6

*La qualité des items d'une épreuve jugée à l'aide des indices de difficulté et
de discrimination*

GUIDE DE L'ANIMATEUR

ACTIVITÉ 6

La qualité des items d'une épreuve jugée à l'aide des indices de difficulté et de discrimination

1. *L'analyse des items*
2. *L'interprétation des indices de difficulté et de discrimination*

1. L'ANALYSE DES ITEMS

Nous portons une grande attention à la rédaction d'une épreuve, aux conditions dans lesquelles elle se passe et aux modalités de correction afin de nous assurer qu'il en résultera une mesure valide et fidèle.

Lorsque les résultats sont comptabilisés, nous pouvons également effectuer une dernière révision afin de vérifier si certaines ambiguïtés ne se seraient pas « gentiment glissées » à l'intérieur de notre épreuve.

Pour nous aider, entre autres, à étudier de plus près la valeur des items d'une épreuve, nous pouvons nous servir des indices de difficulté et de discrimination.

Indice de difficulté :

« Nombre indiquant le degré de difficulté d'un item¹. »

L'indice de difficulté, symbolisé par « P_i », nous renseigne sur la proportion des sujets qui ont réussi l'item étudié. Pour le calculer, il suffit d'établir la proportion des sujets ayant obtenu la bonne réponse par rapport au nombre total de sujets.

$$P_i = \frac{\text{Nombre de sujets ayant la bonne réponse}}{\text{Nombre total de sujets devant répondre à cet item.}}$$

L'indice de difficulté peut prendre toutes les valeurs comprises entre 0,00 et 1,00.

Si $P_i = 1,00$, cela signifie que tous les élèves ayant répondu à l'item 1 ont obtenu la bonne réponse. Nous pouvons donc affirmer que l'item 1 a été « facilement réussi » par les élèves.

1. MEQ, *Éléments de docimologie, Lexique*, fascicule 2, 1985, p. 12.

Si $P_2 = 0,00$, cela signifie que tous les élèves ayant répondu à l'item 2 ont obtenu une mauvaise réponse. Nous pouvons donc constater que l'item 2 a été manqué par la totalité des élèves.

P

0,00 _____ 1,00

<— difficilement réussi facilement réussi —>

Pour les items à réponse construite qui ne sont pas corrigés de façon dichotomique, l'indice de difficulté est obtenu en divisant la moyenne des points obtenus à l'item par le maximum des points accordés.

$$P_3 = \frac{\text{Moyenne des points obtenus à l'item}}{\text{Maximum des points accordés pour l'item}}$$

Si l'item 3 vaut 10 points et que les élèves ont obtenu comme résultat 5, 6, 7, 8, 9 et 10, la moyenne des points obtenus à l'item sera de 7,5 et nous obtiendrons l'indice de difficulté suivant :

$$P_3 = \frac{7,5}{10} = 0,75$$

L'item 3 a donc 0,75 comme indice de difficulté.

Indice de discrimination :

«Nombre indiquant le pouvoir qu'a un item de différencier les élèves forts des élèves faibles tels que départagés par les résultats obtenus à l'ensemble des items d'un instrument de mesure».

L'indice de discrimination, symbolisé par «Q», nous renseigne sur le pouvoir discriminant de l'item étudié. Autrement dit, l'indice de discrimination indique dans quelle mesure l'item différencie les élèves qui ont très bien réussi des élèves qui ont moins bien réussi l'épreuve. Pour le calculer, il suffit d'établir la proportion des sujets ayant obtenu la bonne réponse parmi ceux qui ont obtenu les résultats les plus faibles, et de la soustraire de la proportion des sujets ayant obtenu la bonne réponse parmi ceux qui ont obtenu les meilleurs résultats à l'épreuve.

Le groupe supérieur est constitué par environ 27 p. 100 des élèves qui ont obtenu les résultats les plus forts et le groupe inférieur est constitué, quant à lui, de 27 p. 100 des élèves qui ont obtenu les résultats les plus faibles à l'épreuve.

$$Q_d = \frac{\begin{array}{l} \text{Nombre de sujets} \\ \text{ayant la bonne} \\ \text{réponse dans le} \\ \text{groupe supérieur} \end{array}}{\begin{array}{l} \text{Nombre de sujets} \\ \text{dans le groupe} \\ \text{supérieur} \end{array}} - \frac{\begin{array}{l} \text{Nombre de sujets} \\ \text{ayant la bonne} \\ \text{réponse dans le} \\ \text{groupe inférieur} \end{array}}{\begin{array}{l} \text{Nombre de sujets} \\ \text{dans le groupe} \\ \text{inférieur} \end{array}}$$

L'indice de discrimination peut prendre toutes les valeurs comprises entre -1,00 et +1,00.

Si $Q_4 = +1,00$, cela signifie que l'item 4 discrimine parfaitement les élèves «forts» des élèves «faibles», autrement dit tous les élèves du groupe supérieur ont réussi l'item 4 tandis qu'aucun élève du groupe inférieur n'a réussi l'item.

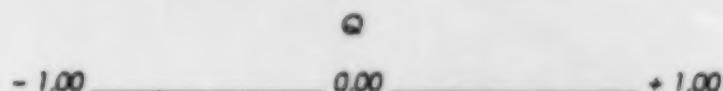
$$Q_4 = \frac{10}{10} - \frac{0}{10} = +1,00$$

Si $Q_5 = 0,00$, cela signifie que l'item 5 ne discrimine pas, c'est-à-dire qu'un même nombre d'élèves dans l'un et l'autre groupe a réussi l'item 5.

$$Q_5 = \frac{7}{10} - \frac{7}{10} = 0,00$$

Si $Q_6 = -1,00$, cela signifie que tous les élèves du groupe inférieur ont réussi l'item 6 tandis qu'aucun élève du groupe supérieur n'a réussi l'item. Un indice de discrimination négatif est habituellement le signe d'un item problématique.

$$Q_6 = \frac{0}{10} - \frac{10}{10} = -1,00$$



← item problématique | item discriminatoire →

item non discriminatoire

La méthode de calcul qui vous est présentée convient bien lorsque l'item à l'étude est à correction objective (dichotomique : succès - échec). Par contre, pour un item à réponse construite qui ne se corrige pas de façon dichotomique, cette méthode ne convient pas. Nous devons recourir à la corrélation afin d'établir la relation entre les résultats à l'item et les résultats à l'épreuve. Si la corrélation tend vers + 1,00, l'item discrimine bien; si la corrélation tend vers 0,00, l'item ne discrimine pas; et si la corrélation est négative, l'item est probablement problématique.

Vous pouvez évaluer le pouvoir discriminant de vos items, sans recourir à la méthode de calcul.

Il suffit, suite à la correction :

1. de classer les copies des élèves du plus haut résultat de l'épreuve au plus bas;
2. de retirer les dix premières copies afin d'établir votre groupe supérieur;
3. de retirer les dix dernières copies afin d'établir votre groupe inférieur;
4. de regarder combien d'élèves du groupe supérieur ont bien répondu à l'item 7 par exemple, et combien d'élèves du groupe inférieur ont également bien répondu à l'item.

Si vous remarquez qu'il y a autant ou plus d'élèves du groupe supérieur, par rapport au groupe inférieur, qui ont bien répondu à l'item 7, vous pouvez «dormir tranquille». Par contre, s'il y a plus d'élèves du groupe inférieur, par rapport au groupe supérieur, qui ont bien répondu à l'item 7, vous devrez relire attentivement l'item, car une ambiguïté s'y est probablement glissée. Si vous êtes incapable de déceler la moindre ambiguïté, demandez à vos élèves, ils vous le diront...

2. L'INTERPRÉTATION DES INDICES DE DIFFICULTÉ ET DE DISCRIMINATION

La relation entre les indices de difficulté et de discrimination.

Si l'item 8 a un indice de difficulté égal à 1,00, quelle sera la valeur de son indice de discrimination?

Réponse : ($Q_2 = 0,00$)

Si l'item 9 a un indice de difficulté égal à 0,00, quelle sera la valeur de son indice de discrimination?

Réponse : ($Q_2 = 0,00$)

Lorsque tous les élèves réussissent un item, on constate que l'item en question ne discrimine aucunement les élèves «forts à l'épreuve» des élèves «faibles à l'épreuve». Il en va de même pour un item qui n'a été réussi par aucun élève.

Un exercice :

Voici les résultats de dix élèves à cinq items. Un point a été accordé pour une bonne réponse et aucun point n'a été accordé pour une mauvaise réponse.

- 1. Calculez le score total de chaque élève.*
- 2. Calculez les indices de difficulté pour chacun des items.*
- 3. Calculez les indices de discrimination pour chacun des items. Formez deux groupes :*
 - groupe supérieur : les trois élèves les plus forts à l'épreuve;*
 - groupe inférieur : les trois élèves les plus faibles à l'épreuve.*

TABLEAU

ITEMS	1	2	3	4	5
ÉLÈVES					SCORE TOTAL
A	1	0	1	1	0
B	0	0	0	1	1
C	1	1	1	1	1
D	1	1	1	1	0
E	0	1	1	1	0
F	1	0	1	1	0
G	1	1	0	1	0
H	0	0	1	1	0
I	1	1	1	1	0
J	0	0	0	0	1
P =					
Q =					

TABLEAU (Solution)

ITEMS	1	2	3	4	5	
ÉLÈVES						SCORE TOTAL
A	1	0	1	1	0	3
B	0	0	0	1	1	2
C	1	1	1	1	1	5
D	1	1	1	1	0	4
E	0	1	1	1	0	3
F	1	0	1	1	0	3
G	1	1	0	1	0	3
H	0	0	1	1	0	2
I	1	1	1	1	0	4
J	0	0	0	0	1	1
P =	0.6	0.5	0.7	0.9	0.3	
Q =	+ 1.0	+ 1.0	+ 0.7	+ 0.3	- 0.3	

L'importance à accorder aux indices de difficulté et de discrimination au moment de l'analyse des items.

Un exercice :

Si vous aviez à améliorer l'épreuve comportant cinq items,

- 1- quel serait l'item à rejeter ou à modifier considérablement? Justifiez votre choix.
- 2- quels seraient les items à conserver? Justifiez votre choix.

Répondez en tenant compte seulement des indices de difficulté et de discrimination que vous venez de calculer.

RÉPONSES

- 1- Item 5, parce que l'indice de discrimination est négatif.
- 2- Le choix des items dépendra du but de l'épreuve. Par contre, il demeure difficile de sélectionner des items sur la seule base de leurs indices.
 - Si nous voulons sélectionner les meilleurs élèves de la classe, les items 1, 2 et 3 sembleraient appropriés parce qu'ils discriminent assez bien.
 - Si le but de l'épreuve est de vérifier l'atteinte des objectifs d'un programme d'études, on ne peut retenir les items en se fiant uniquement aux indices. **Dans ce contexte, pour vérifier la qualité d'un item, il faudra avant tout vérifier la congruence de l'item avec son objectif et son habileté.**

Exemples :

Item 4 ($P = 0,9$)
 ($Q = 0,3$)

Si la congruence est établie entre l'item et l'objectif qu'il mesure, on pourra donc dire que la majorité des élèves a atteint l'objectif. Dans ce contexte, on ne peut éliminer l'item parce qu'il a été réussi facilement par la majorité des élèves. Conséquemment, on ne peut éliminer l'item parce qu'il discrimine peu.

Item 1 ($P = 0,6$)
 ($Q = 1,0$)

Item 2 ($P = 0,5$)
 ($Q = 1,0$)

Pour les items 1 et 2, on ne peut décider d'éliminer ou de retenir ces items en raison de leurs indices de discrimination. Le critère de congruence sera le plus important à considérer pour juger de la qualité de ces items.

En résumé, lorsque nous rédigeons des épreuves qui ont pour but de vérifier l'atteinte des objectifs d'un programme d'études, seul un indice de discrimination **négligé** peut être suffisant pour éliminer ou reformuler un item.

Si vous aviez à rédiger des épreuves de **sélection**, les indices de difficulté et de discrimination vous seraient d'une plus grande utilité.

ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

ACTIVITÉ 6

La qualité des items d'une épreuve jugée à l'aide des indices de difficulté et
de discrimination

GUIDE DU PARTICIPANT

ACTIVITÉ 6

La qualité des items d'une épreuve jugée à l'aide des indices de difficulté et de discrimination

1. L'analyse des items
2. L'interprétation des indices de difficulté et de discrimination

1. L'ANALYSE DES ITEMS

Nous portons une grande attention à la rédaction d'une épreuve, aux conditions dans lesquelles elle se passe et aux modalités de correction afin de nous assurer qu'il en résultera une mesure valide et fidèle.

Lorsque les résultats sont comptabilisés, nous pouvons également effectuer une dernière révision afin de vérifier si certaines ambiguïtés ne se seraient pas «gentiment glissées» à l'intérieur de notre épreuve.

Pour nous aider, entre autres, à étudier de plus près la valeur des items d'une épreuve, nous pouvons nous servir des indices de difficulté et de discrimination.

Indice de difficulté :

«Nombre indiquant le degré de difficulté d'un item¹»

L'indice de difficulté, symbolisé par «P», nous renseigne sur la proportion des sujets qui ont réussi l'item étudié. Pour le calculer, il suffit d'établir la proportion des sujets ayant obtenu la bonne réponse par rapport au nombre total de sujets.

$$P_i = \frac{\text{Nombre de sujets ayant la bonne réponse}}{\text{Nombre total de sujets devant répondre à cet item.}}$$

L'indice de difficulté peut prendre toutes les valeurs comprises entre 0,00 et 1,00.

Si $P_i = 1,00$, cela signifie que tous les élèves ayant répondu à l'item 1 ont obtenu la bonne réponse. Nous pouvons donc affirmer que l'item 1 a été «facilement réussi» par les élèves.

1. MEQ. *Éléments de docimologie, Lexique*, fascicule 2, 1985, p. 12.

Si $P_2 = 0,00$, cela signifie que tous les élèves ayant répondu à l'item 2 ont obtenu une mauvaise réponse. Nous pouvons donc constater que l'item 2 a été manqué par la totalité des élèves.

P

0,00 _____ 1,00

← difficilement réussi facilement réussi →

Pour les items à réponse construite qui ne sont pas corrigés de façon dichotomique, l'indice de difficulté est obtenu en divisant la moyenne des points obtenus à l'item par le maximum des points accordés.

$$P_i = \frac{\text{Moyenne des points obtenus à l'item}}{\text{Maximum des points accordés pour l'item}}$$

Si l'item 3 vaut 10 points et que les élèves ont obtenu comme résultat 5, 6, 7, 8, 9 et 10, la moyenne des points obtenus à l'item sera de 7,5 et nous obtiendrons l'indice de difficulté suivant :

$$P_3 = \frac{7,5}{10} = 0,75$$

L'item 3 a donc 0,75 comme indice de difficulté.

Indice de discrimination :

«Nombre indiquant le pouvoir qu'a un item de différencier les élèves forts des élèves faibles tels que départagés par les résultats obtenus à l'ensemble des items d'un instrument de mesure².»

L'indice de discrimination, symbolisé par «Q», nous renseigne sur le pouvoir discriminant de l'item étudié. Autrement dit, l'indice de discrimination indique dans quelle mesure l'item différencie les élèves qui ont très bien réussi des élèves qui ont moins bien réussi l'épreuve. Pour le calculer, il suffit d'établir la proportion des sujets ayant obtenu la bonne réponse parmi ceux qui ont obtenu les résultats les plus faibles, et de la soustraire de la proportion des sujets ayant obtenu la bonne réponse, parmi ceux qui ont obtenu les meilleurs résultats à l'épreuve.

Le groupe supérieur est constitué par environ 27 p. 100 des élèves qui ont obtenu les résultats les plus forts et le groupe inférieur est constitué, quant à lui, de 27 p. 100 des élèves qui ont obtenu les résultats les plus faibles à l'épreuve.

$$Q_d = \frac{\text{Nombre de sujets ayant la bonne réponse dans le groupe supérieur}}{\text{Nombre de sujets dans le groupe supérieur}} - \frac{\text{Nombre de sujets ayant la bonne réponse dans le groupe inférieur}}{\text{Nombre de sujets dans le groupe inférieur}}$$

L'indice de discrimination peut prendre toutes les valeurs comprises entre -1,00 et +1,00.

2. MEQ. *Éléments de docimologie, Lexique*, fascicule 2, 1985, p. 12.

Si $Q_4 = +1,00$, cela signifie que l'item 4 discrimine parfaitement les élèves «forts» des élèves «faibles», autrement dit tous les élèves du groupe supérieur ont réussi l'item 4 tandis qu'aucun élève du groupe inférieur n'a réussi l'item.

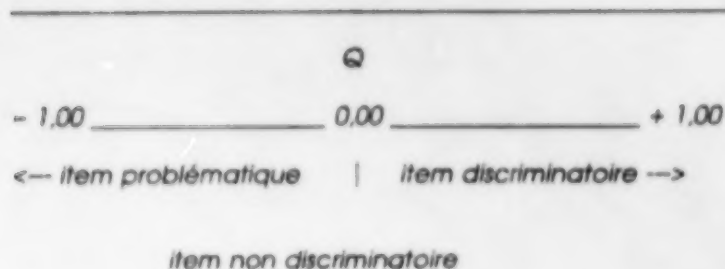
$$Q_4 = \frac{10}{10} - \frac{0}{10} = +1,00$$

Si $Q_5 = 0,00$, cela signifie que l'item 5 ne discrimine pas, c'est-à-dire qu'un même nombre d'élèves dans l'un et l'autre groupe a réussi l'item 5.

$$Q_5 = \frac{7}{10} - \frac{7}{10} = 0,00$$

Si $Q_6 = -1,00$, cela signifie que tous les élèves du groupe inférieur ont réussi l'item 6 tandis qu'aucun élève du groupe supérieur n'a réussi l'item. Un indice de discrimination négatif est habituellement le signe d'un item problématique.

$$Q_6 = \frac{0}{10} - \frac{10}{10} = -1,00$$



La méthode de calcul qui vous est présentée convient bien lorsque l'item à l'étude est à correction objective (dichotomique : succès - échec). Par contre, pour un item à réponse construite qui ne se corrige pas de façon dichotomique, cette méthode ne convient pas. Nous devons recourir à la corrélation afin d'établir la relation entre les résultats à l'item et les résultats à l'épreuve. Si la corrélation tend vers + 1,00, l'item discrimine bien; si la corrélation tend vers 0,00, l'item ne discrimine pas; et si la corrélation est négative, l'item est probablement problématique.

Vous pouvez évaluer le pouvoir discriminant de vos items, sans recourir à la méthode de calcul.

Il suffit, suite à la correction :

1. de classer les copies des élèves du plus haut résultat de l'épreuve au plus bas;
2. de retirer les dix premières copies afin d'établir votre groupe supérieur;
3. de retirer les dix dernières copies afin d'établir votre groupe inférieur;
4. de regarder combien d'élèves du groupe supérieur ont bien répondu à l'item 7 par exemple, et combien d'élèves du groupe inférieur ont également bien répondu à l'item.

Si vous remarquez qu'il y a autant ou plus d'élèves du groupe supérieur, par rapport au groupe inférieur, qui ont bien répondu à l'item 7, vous pouvez «dormir tranquille». Par contre, s'il y a plus d'élèves du groupe inférieur, par rapport au groupe supérieur, qui ont bien répondu à l'item 7, vous devrez relire attentivement l'item, car une ambiguïté s'y est probablement glissée. Si vous êtes incapable de déceler la moindre ambiguïté, demandez à vos élèves, ils vous le diront...

2. L'INTERPRÉTATION DES INDICES DE DIFFICULTÉ ET DE DISCRIMINATION

La relation entre les indices de difficulté et de discrimination.

Si l'item 8 a un indice de difficulté égal à 1,00, quelle sera la valeur de son indice de discrimination?

Réponse : ($Q_8 = 0,00$)

Si l'item 9 a un indice de difficulté égal à 0,00, quelle sera la valeur de son indice de discrimination?

Réponse : ($Q_9 = 0,00$)

Lorsque tous les élèves réussissent un item, on constate que l'item en question ne discrimine aucunement les élèves «forts à l'épreuve» des élèves «faibles à l'épreuve». Il en va de même pour un item qui n'a été réussi par aucun élève.

Un exercice :

Voici les résultats de dix élèves à cinq items. Un point a été accordé pour une bonne réponse et aucun point n'a été accordé pour une mauvaise réponse.

1. Calculez le score total de chaque élève.
2. Calculez les indices de difficulté pour chacun des items.
3. Calculez les indices de discrimination pour chacun des items. Formez deux groupes :
 - groupe supérieur : les trois élèves les plus forts à l'épreuve;
 - groupe inférieur : les trois élèves les plus faibles à l'épreuve.

TABLEAU

ITEMS	1	2	3	4	5
ÉLÈVES					SCORE TOTAL
A	1	0	1	1	0
B	0	0	0	1	1
C	1	1	1	1	1
D	1	1	1	1	0
E	0	1	1	1	0
F	1	0	1	1	0
G	1	1	0	1	0
H	0	0	1	1	0
I	1	1	1	1	0
J	0	0	0	0	1

P =

Q =

L'importance à accorder aux indices de difficulté et de discrimination au moment de l'analyse des items.

Un exercice :

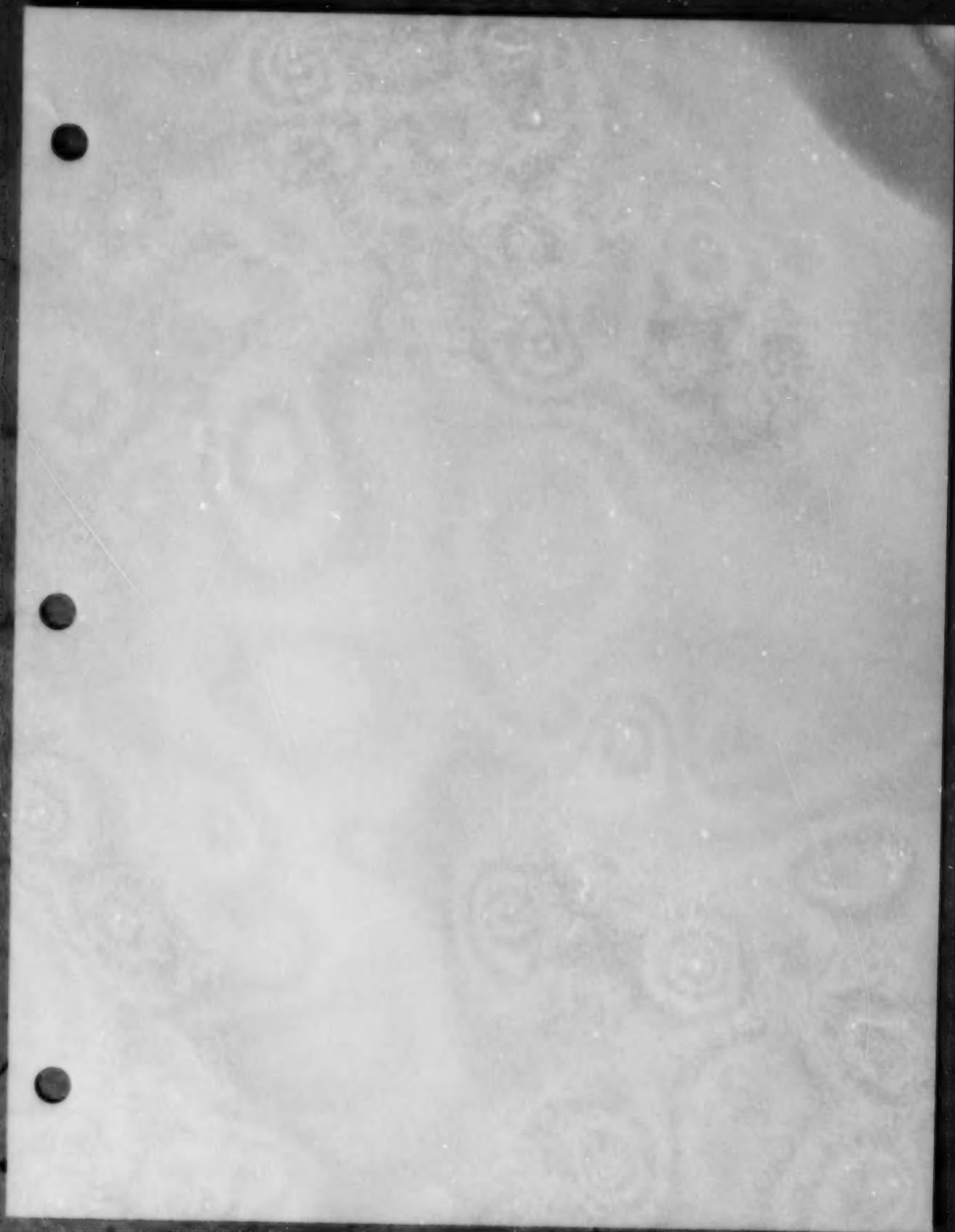
Si vous aviez à améliorer l'épreuve comportant cinq items,

- 1- quel serait l'item à rejeter ou à modifier considérablement? Justifiez votre choix.
- 2- quels seraient les items à conserver? Justifiez votre choix.

Répondez en tenant compte seulement des indices de difficulté et de discrimination que vous venez de calculer.

En résumé, lorsque nous rédigeons des épreuves qui ont pour but de vérifier l'atteinte des objectifs d'un programme d'études, seul un indice de discrimination **négalif** peut être suffisant pour éliminer ou reformuler un item.

Si vous aviez à rédiger des épreuves de **sélection**, les indices de difficulté et de discrimination vous seraient d'une plus grande utilité.



DOCUMENT DE BASE

ÉVALUATION SOMMATIVE DES SCIENCES AU SECONDAIRE
À PARTIR DU MODÈLE RETENU DANS LES DÉFINITIONS DU DOMAINE

DOCUMENT DE BASE

LES DÉFINITIONS DU DOMAINE

TABLE DES MATIÈRES

1.	Historique	p. 3
2.	L'évaluation sommative	p. 7
3.	Les qualités des instruments de mesure	p. 8
4.	La définition du domaine	p. 9
5.	Les habiletés	p. 11
6.	Les dimensions	p. 12
Liste des ouvrages et du matériel utilisés dans la préparation du document		p. 13

1. Historique

La recherche en mesure et évaluation

Vous êtes-vous déjà demandé pourquoi la mesure et l'évaluation des apprentissages sont tellement à la mode depuis quelques années? Pourquoi se fait-on inonder de nouvelles techniques et de nouveaux termes? C'est pourtant depuis des siècles que le maître évalue son élève...

Mais il faut attendre au début du 20^e siècle pour voir paraître le premier ouvrage traitant de l'évaluation en éducation. Cette œuvre, nous la devons à celui que l'on a surnommé "father of modern educational measurements" : Edward L. Thorndike.

On reconnaît maintenant l'évaluation des apprentissages comme un champ d'application des théories de la mesure. C'est donc au cours du présent siècle que la recherche dans ce domaine s'est intensifiée. Mais il faut encore attendre au milieu du 20^e siècle pour voir paraître certains résultats de ces recherches dans nos pratiques évaluatives.

Quelques développements

Voici donc très simplement quelques développements qui ont très certainement préparé l'avènement des définitions du domaine.

On reconnaît la qualité d'un instrument de mesure par sa validité et sa fidélité. L'importance de ces concepts est universellement reconnue en recherche et ce, que le champ soit les sciences pures, les sciences appliquées ou les sciences humaines. Ce qui peut sembler nouveau, c'est le fait que les concepts de validité et de fidélité soient associés à des instruments de mesure de l'apprentissage. Effectivement, dans son sens large, l'instrument servant à mesurer des apprentissages doit posséder les mêmes qualités qu'un instrument servant à mesurer la longueur d'une table ou la densité d'un métal.

Validité

Dans le contexte de la mesure des apprentissages en éducation, la validité signifie la «qualité d'un instrument qui mesure bien tout ce qu'il prétend mesurer, et rien de

plus¹.» Autrement dit, l'ensemble des items de votre épreuve représente-t-il bien l'ensemble des objectifs que vous vous proposez de mesurer et rien de plus? Si oui, votre épreuve est fidèle.

Mais comment pouvez-vous vous en assurer?

- 1) En déterminant clairement les objectifs dont on veut mesurer l'atteinte. On peut le faire sous forme de liste ou sous forme de tableau de spécifications.

Un tableau de spécifications n'est en fait qu'un tableau à double entrée où l'on trouve d'une part les objectifs ou les notions d'un programme d'études et d'autre part les niveaux taxinomiques. Vous avez sans doute remarqué que cette notion de tableau de spécification, s'apparente étroitement à la notion de définition du domaine.

- 2) En vérifiant la congruence entre l'objectif dont on veut mesurer l'atteinte et le contenu de l'item.

Fidélité

Toujours dans le contexte de la mesure des apprentissages en éducation, la deuxième qualité recherchée est la fidélité. Elle signifie «qualité d'un instrument de mesure dont les résultats demeurent stables pour les mêmes sujets qui se trouvent dans des conditions identiques lors d'administrations différentes².» Autrement dit, le résultat que vous avez donné à Justine ressemble-t-il au rendement de Justine?

Si oui, votre épreuve est fidèle.

1. MINISTÈRE DE L'ÉDUCATION, *Éléments de docimologie, lexique*, Québec, 1985, p. 17.

2. *Idem*, p. 11.

Mais comment pouvez-vous vous en assurer?

- 1) En portant attention aux erreurs inhérentes à l'épreuve : la longueur, le vocabulaire, la limite de temps, la correction, la formulation des items, etc.
- 2) En portant attention aux erreurs dues à la passation de l'épreuve : conditions d'administration, etc.

Validité et
fidélité

Les concepts de validité et de fidélité sont liés mathématiquement. De plus, il faut se rappeler que la validité est la qualité la plus importante. Concrètement, si l'on se préoccupe de la validité de notre épreuve, la fidélité de cette même épreuve s'en trouvera automatiquement augmentée.

On peut donc comprendre toute l'importance que revêt la qualité de la validité lorsque nous préparons des épreuves sommatives. Rappelons que la définition du domaine renferme les renseignements nécessaires afin d'obtenir une épreuve de qualité.

Le contexte
québécois

En 1980, la Direction du développement en évaluation (DDE) adopte le concept de définition du domaine au moment de la préparation d'une banque d'instruments de mesure connue aujourd'hui sous le nom de BIM.

La définition du domaine a donc été décrite comme un moyen efficace d'obtenir une synthèse des éléments essentiels d'un programme d'études pour des fins d'évaluation sommative.

Il faut se rappeler que la définition du domaine vise essentiellement à assurer la congruence entre un programme d'études et tout instrument de mesure devant servir à une évaluation sommative.

En 1985, on voit paraître les ébauches de la définition du domaine du programme d'études Histoire du Québec et du Canada en 4^e secondaire.

En 1986, on poursuit la production de définitions du domaine pour certains programmes d'études, tant au primaire qu'au secondaire.

Avant que les responsables des programmes d'études définissent les habiletés propres à chaque programme à l'intérieur des définitions du domaine, on avait recours à différentes taxinomies. Celle de Bloom était sans doute la plus connue et la plus utilisée. Pour certains programmes, la taxinomie de Bloom convenait très bien, mais pour d'autres, elle convenait moins bien. C'est pour cette raison qu'il y a eu une volonté de recourir à des habiletés propres à chacun des programmes d'études.

2. L'évaluation sommative

Évaluation sommative

«Démarche qui vise à porter un jugement sur le degré de réalisation des apprentissages visés par un programme ou une partie terminale de programme ou encore d'un ensemble équivalent d'apprentissages faits en dehors de l'école, en se basant sur des données pertinentes, pour prendre des décisions relatives :

- au passage à la classe supérieure,
- à la sanction des études,
- à la reconnaissance des acquis expérimentiels (équivalences),
- à l'orientation de l'élève³»

3. MINISTÈRE DE L'ÉDUCATION, *Éléments de docimologie, lexique*, Québec, 1985, p. 10.

3. Les qualités des instruments de mesure

Validité

Qualité la plus importante.

L'ensemble des items de votre épreuve représente-t-il bien l'ensemble des objectifs que vous voulez mesurer? Si oui, votre épreuve est valide.

Les moyens de s'assurer de la validité d'une épreuve :

- déterminer clairement les objectifs que l'on veut mesurer;
- se faire un tableau de spécifications (définition du domaine);
- voir à ce que les items soient bien construits;
- s'assurer de la congruence entre l'objectif dont on veut mesurer l'atteinte et le contenu des items.

Fidélité

Les résultats de l'épreuve sont-ils stables?
«Est-ce que le résultat de Justine ressemble à Justine?»

Erreur de mesure (E) faible.

Les moyens de s'assurer de la fidélité d'une épreuve :

- porter attention aux erreurs inhérentes à l'épreuve;

longueur de l'épreuve,
vocabulaire utilisé,
ambiguïté de l'épreuve,
temps limité,
correction, etc.
- porter attention aux erreurs dues à la passation de l'épreuve;

conditions de passation, etc.
- porter attention aux erreurs provenant de l'élève lui-même (les plus difficiles à contrôler).

4. **La définition du domaine**

Domaine

«Univers d'éléments auquel un programme, un objectif, une épreuve ou un item fait référence⁴.»

D'une façon plus particulière, le domaine est la «description et le classement des différents éléments essentiels et représentatifs d'un programme ou d'un objectif aux fins de l'évaluation. (...)

Le domaine délimite ce sur quoi doit porter l'examen utilisé pour appuyer son jugement; il permet d'établir ou de vérifier le lien entre le programme ou l'objectif et l'examen en question; finalement, il sert de base à l'interprétation des résultats. De plus, un domaine bien défini identifie tous les paramètres à considérer et contient toutes les conditions d'attribution des unités ou des reconnaissances d'acquis, lorsqu'il s'agit de sanction d'études⁵.»

C'est en fait un tableau à double entrées.

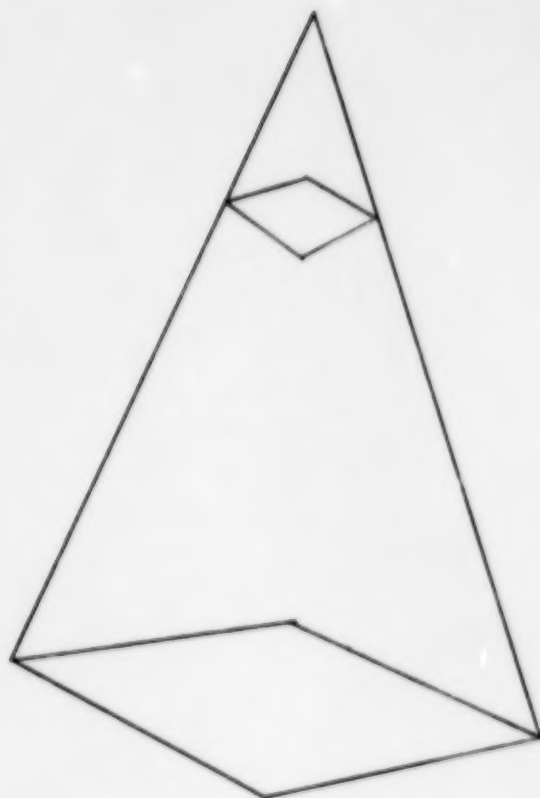
ACÉTATE 1

4. MINISTÈRE DE L'ÉDUCATION, *Éléments de docimologie, lexicque*, Québec, 1985, p. 9.

5. BIM, *Guide pour la construction d'instruments de mesure dans le cadre de la banque d'instruments de mesure*, chapitres II - III, 1984, p. 2-2.

ÉVALUER

B-1



DÉFINITION
DU
DOMAINE

PROGRAMME D'ÉTUDES
(APPRENTISSAGES)

5. Les habiletés

Habileté

«Ensemble de comportements de l'ordre du savoir-faire, mis en oeuvre dans une situation donnée. Il peut s'agir de savoir-faire cognitifs, affectifs ou psychomoteurs⁶.»

Une habileté telle que décrite dans les définitions du domaine est le scénario le plus probable, relativement à la tâche à accomplir, pour la majorité des élèves. On pourrait parler d'une habileté comme d'une «dominante» ou d'une «majeure».

Trois des quatre habiletés ne sont pas exclusives en soi. Elles représentent l'apprentissage véhiculé par le programme d'études en sciences physiques. Cette limite aura des répercussions sur l'exclusivité des dimensions.

6. MINISTÈRE DE L'ÉDUCATION, *Éléments de docimologie, lexique*, Québec, 1985, p. 11.

6. Les dimensions

Dimension

Chaque regroupement obtenu en combinant un thème et une habileté représente une dimension.

L'ensemble des dimensions représente le domaine.

Les dimensions ne sont pas exclusives en soi.

LISTE DES OUVRAGES ET DU MATÉRIEL UTILISÉS DANS LA PRÉPARATION DU DOCUMENT

ARSENAULT, LOUISE. **Sessions de formation en évaluation des apprentissages : Documents d'accompagnement**, 1986-1991.

AUGER, RÉJEAN. **La définition de domaine**, Document de travail.

BERNIER, JEAN-JACQUES. **Théorie des tests**, Principes et techniques de base, deuxième édition, Gaétan Morin éditeur, 1985.

BIM. **Guide pour la construction d'instruments de mesure dans le cadre de la banque d'instruments de mesure**, chapitres II - III, IV - V «s.éd.», 1984.

CHARLES, RANDALL, LESTER, FRANK, O'DAFFER, PHARES. **How to evaluate progress in problem solving**, The National Council Of Teachers Of Mathematics Inc., 1987.

DE LANDSHEERE, GILBERT. **Dictionnaire de l'évaluation et de la recherche en éducation**, Presses universitaires de France, 1979.

EBEL, ROBERT L. **Essentials of educational measurement**, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1972.

LUCIER CHARLES, DENISE et al. **Définition de domaine, rapport de comité**, Montréal, Mars 1988.

MINISTÈRE DE L'ÉDUCATION. **Cadre général d'élaboration des épreuves uniques**, Direction générale de l'évaluation, Québec, 1986.

MINISTÈRE DE L'ÉDUCATION. **Éléments de docimologie**, fascicules 1, 2, 3 et 4, Québec, Direction générale des publications gouvernementales du ministère des Communications, Québec, 1985-1987.

MINISTÈRE DE L'ÉDUCATION. **Guide pédagogique, Primaire, Mathématique**, Fascicule K, Résolution de problèmes, Orientation générale, Document 16-2300-11, Direction générale des programmes, Québec, 1988.

MINISTÈRE DE L'ÉDUCATION. **MATHÉMATIQUE, Recueil d'activités en vue d'un perfectionnement en évaluation sommative, Primaire et secondaire**, Direction du développement de l'évaluation, Québec, tome 1 : 1989, tome 2 : 1990.

MINISTÈRE DE L'ÉDUCATION. *Politique générale d'évaluation pédagogique, secteur du préscolaire, du primaire et du secondaire*, Direction générale du développement pédagogique, Québec, 1981.

SCALLON, GÉRARD. *L'évaluation formative des apprentissages, l'instrumentation*, Les presses de l'Université Laval, 1988.

SCALLON, GÉRARD. *L'évaluation formative des apprentissages, notes de cours*.

FORMULES D'ÉVALUATION DE L'ATELIER

ÉVALUATION DE L'ATELIER

TA	tout à fait en accord
A	en accord
D	en désaccord
TD	tout à fait en désaccord

Directive : encerclez la réponse de votre choix.

- | | | | | | |
|-----|---|----|---|---|----|
| 1- | Je suivrais un autre atelier présenté de cette manière. | TA | A | D | TD |
| 2- | Il était facile de demeurer attentif. | TA | A | D | TD |
| 3- | L'animatrice a fait la synthèse de la matière. | TA | A | D | TD |
| 4- | L'animatrice favorisait l'expression de points de vue nouveaux. | TA | A | D | TD |
| 5- | Certains concepts n'étaient pas tellement bien expliqués. | TA | A | D | TD |
| 6- | L'animatrice connaissait bien le contenu. | TA | A | D | TD |
| 7- | Quelquefois, je n'étais pas très intéressé par le contenu de l'atelier. | TA | A | D | TD |
| 8- | Le contenu de cet atelier va m'aider dans mon enseignement. | TA | A | D | TD |
| 9- | Les documents reçus me seront très utiles. | TA | A | D | TD |
| 10- | L'atmosphère a été vivante tout au long de l'atelier. | TA | A | D | TD |
| 11- | J'encouragerais mes collègues à participer à des ateliers de ce genre. | TA | A | D | TD |
| 12- | Il devrait y avoir une suite à cet atelier. | TA | A | D | TD |

Merci et soyez assurés que l'information que vous m'avez transmise recevra une grande attention.

ÉVALUATION DE L'ATELIER

TA	tout à fait en accord
A	en accord
D	en désaccord
TD	tout à fait en désaccord

Directive : encerclez la réponse de votre choix.

- | | | | | | |
|-----|---|----|---|---|----|
| 1- | je suivrais un autre atelier présenté de cette manière. | TA | A | D | TD |
| 2- | Il était facile de demeurer attentif. | TA | A | D | TD |
| 3- | L'animateur a fait la synthèse de la matière. | TA | A | D | TD |
| 4- | L'animateur favorisait l'expression de points de vue nouveaux. | TA | A | D | TD |
| 5- | Certains concepts n'étaient pas tellement bien expliqués. | TA | A | D | TD |
| 6- | L'animateur connaissait bien le contenu. | TA | A | D | TD |
| 7- | Quelquefois, je n'étais pas très intéressé par le contenu de l'atelier. | TA | A | D | TD |
| 8- | Le contenu de cet atelier va m'aider dans mon enseignement. | TA | A | D | TD |
| 9- | Les documents reçus me seront très utiles. | TA | A | D | TD |
| 10- | L'atmosphère a été vivante tout au long de l'atelier. | TA | A | D | TD |
| 11- | J'encouragerais mes collègues à participer à des ateliers de ce genre. | TA | A | D | TD |
| 12- | Il devrait y avoir une suite à cet atelier. | TA | A | D | TD |

Merci et soyez assurés que l'information que vous m'avez transmise recevra une grande attention.



Gouvernement du Québec
Ministère
de l'éducation

16-7560